The Impact of Blockchain Technology on Clinical Trial Data Integrity

By

Karl Cullen

A thesis submitted in partial fulfilment for the requirements for MSc in Pharmaceutical Business & Technology

Innopharma Faculty of Pharmaceutical Science

Griffith College, Dublin

August 2020

Declaration

I declare that the work in this dissertation titled “The Impact of Blockchain Technology on Clinical Data Integrity” which I now submit for assessment as part of the partial fulfilment for the Masters of “Pharmaceutical Business and Technology” is the result of my own work and that I have referenced the work of others and due acknowledgement is given.

Candidate Signature: 

Date: 28/08/2020

Supervisor Name: Kathy Clarke

Supervisor Signature: 

Date: 28/08/2020

Acknowledgments

I would like to firstly acknowledge and thank Kathy Clarke for her invaluable support and guidance throughout this research study. Your indispensable insight and continuous encouragement pushed me to have faith and pride in my work. I would also like to thank Finbarr Sheehy for assisting me in establishing the foundation of this research. Finally, I would like to thank all the participants who took time to participate in this research, your participation is deeply appreciated.

Table of Contents

List of Figures

List of Abbreviations

BC - Blockchain

BCT – Blockchain Technology

EMA – European Medicines Agency

FDA – Food and Drug Administration

GCP – Good Clinical Practice

GDPR – General Data Protection Regulation

MHRA - Medicines and Healthcare products Regulatory Agency

P2P – Peer-2-Pee**r**

POC – Proof-of-Concept

SDV – Source Data Verification

The Impact of Blockchain Technology on Clinical Trial Data Integrity

Abstract

Karl Cullen

In this study it was shown how a new and emerging technology known as Blockchain, a data management platform, possesses the potential to address some of the predominant issues pertaining to clinical data integrity within clinical trial research studies. However, many obstacles and challenges persist with its implementation and these must be addressed, understood and solved before the technology can be utilized to manage clinical data. The main research objective of this study was to examine how Blockchain could be suited to address the concerns relating to clinical data integrity and to address the perceptions of participants involved with clinical data management to the technology and its potential application within clinical trials. Through an exploratory quantitative and qualitative mixed methods approach aligned with a pragmatist research philosophy, primary data was collected from participants who were purposively selected from a homogenous sampling pool that consisted of those with backgrounds in clinical data management. Primary data produced from this study highlighted the correlation of data issues that contribute towards the lack of integrity of clinical data with those found in secondary research. It was also found that the perception of Blockchain was generally positive and favorable towards its adoption within clinical trials. However, there was a degree of uncertainty surrounding the security of data within Blockchain. Various challenges were identified, primarily, the complexity of the technology and challenges related to compliancy with data protection laws. The conclusive determination of Blockchain's suitability to address clinical data concerns, however, remains unanswered and further research is needed to test the application of Blockchain in real-world clinical trial environments. Further research and studies are also necessary to address the many challenges that face Blockchains implementation with clinical data management.

Chapter 1

Introduction

1.1 Introduction

Information technology has rapidly been evolving in recent years to coincide with the increasing production and complexity of data. Data is being produced in such large quantities that new technologies must be developed that possess superior capabilities for the managing, analyzing, and storing of large complex data streams (Bashir, 2017). One area in which data is growing exponentially is clinical trials (Mackey *et al.,* 2019). Clinical trials are intricate, time-consuming and expensive processes that produce immense quantities of complex and medically relevant data (Shamley and Wright, 2017). The World Health Organization (WHO) describe clinical trials as "a type of research that studies new tests and treatments and evaluates their effects on human health outcomes*"* (WHO, 2018). The number of clinical trials have been growing in recent years and consequently so too has the quantity and complexity of the data that they produce (Shamley and Wright, 2017).

The Food and Drug Administration (FDA) and the Medicines and Healthcare products Regulatory Agency (MHRA) held a joint Good Clinical Practice (GCP) workshop in 2018 to discuss the increasing need for clinical data improvement processes due to the increasing numbers of trials, the increasing number of sites per trial, and the limited resources for oversight bodies in regulating and monitoring clinical trials across multiple locations (Khin *et al.,* 2020). A concluding remark from these discussions were that "the importance of data integrity in clinical trials cannot be overstated" (Khin *et al.,* 2020). Segen's medical dictionary defines data integrity as "a dimension of data contributing to trustworthiness and pertaining to the systems and processes for data capture, correction, maintenance, transmission and retention" (Segen's Medical Dictionary, 2012). Additionally, the FDA issued a guidance document in 2016 titled "Data Integrity and Compliance" which defines data integrity as: "the completeness, consistency, and accuracy of data. Complete, consistent, and accurate data should be attributable, legible, contemporaneously recorded, original or a true copy, and accurate" (FDA, 2016).

Data integrity is a fundamental requirement within CT research as all produced clinical data should represent a true and accurate reflection of the methods, outcomes and results of a trial, regardless of whether the data supports, disproves or challenges the trials original hypothesis (Friar and Kirsh, 2019). The clinical data functions to justify the efficacy and safety of a new medical intervention, treatment or drug, and is therefore critically important for the safety, health, and welfare of prospective patients (Shamley and Wright, 2017). How data is managed, stored and shared has been cited as being a recurring issue within clinical research and has been the subject of ongoing continuous improvement endeavors by regulatory authorities in recent years (Khin *et al.,* 2020). Concerns of data transparency and integrity are cited to be key issues in clinical trials due to data inconsistencies, erroneous data, and data misconduct (George and Buyse, 2015; George, 2016; Khin *et al.,* 2020). The Medical Research Council (MRC) defines research

data misconduct as the "fabrication, falsification, plagiarism, misrepresentation, mismanagement or inadequate preservation of data" (MRC, 2014). As the number of clinical trials and associated complexity and monetary investments grow, there may be a need to implement new data management systems that offer true data transparency, provenance and integrity (Hölbl *et al.*, 2018; Maslove *et al.*, 2018).

Blockchain technology (BCT) is a relatively nascent technology that was originally designed for the publicly perceived controversial cryptocurrency known as Bitcoin in 2008 (Nakamoto, 2009; Bashir, 2017). Blockchains (BC) are decentralized public (or private) ledgers that utilize intelligent cryptographic hashing functions and sequential timestamping of transactional information to provide consecutively chained blocks of information in an append only, fully transparent, and tamper-proof ledger, that is spread across a peer-to-peer network (P2P) (Zheng *et al.*, 2017). Following its original application with Bitcoin, Blockchains innate technological potential data solutions were quickly realized, and the technology rapidly began integrating into the finance industry. BCT has been cited as being a revolutionary and potentially "*disruptive*" technology (Mackey *et al.*, 2019) and in recent years has begun diffusing beyond cryptocurrency and finance into sectors such as life sciences, healthcare, food and pharmaceutical supply chains and government. The life sciences and healthcare industries are where BCT can have a far-reaching impact as these areas rely heavily on the transparency, traceability and trustworthiness of data shared amongst a network of stakeholders such as patients, healthcare providers, insurers, pharmaceutical companies and clinical researchers (Holdowsky, Lele and Lougheed, 2018). In 2016 there were 5 PubMed publications with the word 'Blockchain' in the title or abstract as compared to 2018 when there were 64 publications. Healthcare data represented 32% of the total publications while clinical trials represented 4% (Mackey *et al.*, 2019). There is a clear growing interest in this technology which is supported by its expected market growth. At present Blockchains market size is valued at $3 billion and is expected to grow to $39.7 billion by 2025 (MarketsandMarkets, 2020). However, whether BC can be adapted for applicatory purposes within clinical trial data management remains to be seen, but there is a clear and evident interest in the proposition.

BC has been proposed as a pragmatic solution for a myriad of data issues seen across industries and its potential benefits have begun to seep into the clinical research industry (Holdowsky, Lele and Lougheed, 2018; Hölbl *et al.*, 2018; Mackey *et al.*, 2019). Blockchain technology (BCT) offers advantages such as (Kuo, Kim and Ohno-Machado, 2017; Hölbl *et al.*, 2018):

- Fully transparent chronological and immutable data trails
- No central intermediary to validate transactions

- Limits the possibility of a single-point-of-failure by being distributed across a multitude of nodes
- Disincentives fraudulent activities through intelligent cryptographic functions
- Auditability and traceability of all data
- Increased security and accessibility of all data

BCT offers a new way to manage and store data that could facilitate innovative methods for improving clinical data integrity (Benchoufi and Ravaud, 2017). There have been numerous discussions, pilots, and proof-of-concept (POC) studies to investigate the viability and efficacy of BCT in addressing some of the challenges associated with CT studies such as data integrity, recruitment and retention, management of informed consents, data provenance and data auditing for regulatory purposes (Benchoufi, Porcher and Ravaud, 2018; Maslove *et al.*, 2018; Wong, Bhattacharya and Butte, 2019). BCT, while having existed since 2008, is still in its infancy of implementation within clinical research and industry (Mackey *et al.*, 2019). There are many challenges that must be overcome, and it is inevitable that many more obstacles will present in the future. Besides technical challenges, there are social considerations to contemplate as is the case with any new technology, especially one that will effectively store and maintain sensitive data in an age where data forms the backbone of the digital society.

1.2 Research Purpose

The focus of this research is to gain a deeper understanding of some of the key challenges associated with the integrity of clinical data and how an emerging technology such as Blockchain could address those concerns from a clinical data researcher's perspective. BCT consists of a complex architectural infrastructure that is beyond the scope of this research, as is the actual technical application of the technology for clinical trials, however this is not the aim of the research. There are numerous ongoing and completed proof-of-concept studies that examine the technical applications of this technology with specific areas within clinical trials (Benchoufi, Porcher and Ravaud, 2018; Maslove *et al.*, 2018; Wong, Bhattacharya and Butte, 2019). However, this research intends to examine the social awareness and attitudes to this technology and to subsequently attempt to determine if BCT could alleviate some of the data integrity issues encountered in clinical research from the perspective of those directly involved in clinical data management.

In order to assess the suitability of Blockchain's application in clinical research, study participants include those working within clinical trials who are involved in clinical data management such as Clinical Data Managers and Clinical Data Analysts who are key study participants as they are responsible for the management of data within a clinical trial and represent a key source of primary insight into the research topic. Literature is explored to determine the extent of the

concerns associated with clinical data integrity and study participants are interview through an online survey to determine if the literature correlates with their real-world experiences. Through a mix of qualitative and quantitative research, specific issues of data integrity will be investigated and theories of Blockchain technology applications are applied in order to determine the potential suitability of its application in improving data integrity.

1.3 Research Objective

The research objectives of this study are as follows;

- To gain a deeper understanding of the some of the key issues surrounding clinical research data integrity
- To evaluate the current knowledge and awareness of Blockchain within the clinical research industry
- To determine the suitability of Blockchain in addressing the key issues surrounding clinical data integrity from a clinical data management perspective through quantitative and qualitative research

As there are no real-world examples of Blockchain being implemented within a clinical research setting at the time of conducting this research, the research here is primarily exploratory in nature, therefore no formal hypothesis can be formulated. With any new technology there are many unknown variable factors and often obstacles for implementation are unforeseen.

1.4 Justification for Study

This research was undertaken due to the growing importance and relevance of data management in modern society and the associated issues relating with data integrity. This diffuses into clinical research data as the accuracy and dependability on the integrity of the data is instrumental for the successful outcome of the trial's treatment; on which many prospective patients may depend upon. Additionally, Blockchain represents a truly potential solution for many of the data concerns seen in modern day data management across a range of industries. Its inherent advantages over current data management systems offer potential solutions that are worth exploring within clinical research. Blockchain is already being implemented and trialed across industries and areas such as the finance industry and the supply chain traceability within the food and pharmaceutical industries. With so much data being produced from clinical trials the coalescence of clinical data management with Blockchain seems like a truly beneficial endeavor to improve the integrity of clinical data. As Blockchain is slowly encroaching into the clinical research industry it is necessary to evaluate the perceptions and attitudes towards the technology as the technology will only prove to be effective if adopted by all stakeholders. It is the intent of this research to explore

those attitudes and also to outline the potential benefits Blockchain holds for the future of clinical data management.

1.5 Structure of the Study

Chapter 1 introduces the dissertation and provide a brief background to the topic of clinical research integrity and a short introduction and explanation to Blockchain technology and clinical trials. The purpose of the research is outlined as well as the research objectives, justification and the structure of the dissertation.

Chapter 2 consists of a literature review of clinical trial data management and some of the key issues surrounding integrity. Some of these issues include clinical data fraud and misconduct, and data traceability and transparency. An inclusion of how Blockchain works is included as this is a relatively new technology and will add context to the overall research. It is necessary to outline the functionalities and basic workings of Blockchain in order to examine its relevancy to clinical research data integrity such as traceability and transparency. Some of the key challenges with Blockchain and clinical trial research are discussed which include incompatibilities with data protection laws. Areas of clinical data integrity such as issues involving clinical data research misconduct and fraud as well as other clinical data concerns are highlighted from the literature which form the basis for the research methodology design. A conclusion of the literature is outlined along with gaps discovered and finally the conceptual framework is outlined.

Chapter 3 outlines the research paradigm and the research strategy that the researcher used. A mixed methods approach was adopted aligning with a pragmatist research philosophy. The research overall is considered exploratory due to the nature of the infancy of the technology within the clinical research industry. An online survey consisting of quantitative and qualitative elements was issued to participants who were purposively chosen from a homogenous sample population. Participants were purposely selected who represent those with professions and backgrounds in clinical data management. Thematic analysis was applied to analyze the qualitative data produced while quantitatively produced data was analyzed in order to provide a complementary analysis with the qualitative data. Utilizing both methods allow for the breadth and depth of the responses to be further expanded.

Chapter 4 evaluates and examines the results gained from the research applied methodologies and links to the research aims and overall topic. Thematic analysis is applied which highlights themes that arose from participants qualitative responses. These themes involve clinical data auditability, quality, transparency as well themes involving patients' roles and vaccinations. Quantitatively produced data highlights some of the attitudes and perceptions to issues of clinical data integrity as well as Blockchain-related perceptions. The two sets of data are outlined in a complementary manner and function to support one another. Finally, a discussion of the findings

is included, and links are made to findings made from the literature review and correlations are highlighted between primary and secondary research data.

Chapter 5 outlines the conclusion, limitations and recommendations of the study. Conclusions made from the research include how Blockchain has a strong suitability for addressing the clinical data concerns highlighted from the primary research, however there are numerous obstacles and challenges the technology must overcome and these are outlined here. Various limitations are identified and addressed such as the limited scope of the adopted research methodology. Finally, recommendations for further research are included and these consist of areas such as identifying and addressing the challenges that Blockchain must overcome before consideration of implementing the technology within clinical research.

Chapter 2
Literature Review

2.1 Topic Outline

This chapter outlines a review of the relevant literature and functions to support the study. The literature review attempts to gain a deeper understanding to some of the predominant data integrity issues that are associated with clinical trials. It also investigates why they may occur, while simultaneously exploring a new potential digital data management platform – Blockchain – and how this technology could address some of the main clinical data concerns that are discovered from the literature. The topics of research outlined in this section include an overview of BCT and how it works. This is necessary to understand its potential role within CT research and introduces the technology to the reader. Clinical data management is then explored in relation to clinical data misconduct and fraud. This links with the objective of gaining a deeper understanding to the issues that pertain to clinical data integrity. Following this a review of completed studies involving the application of BC within clinical research is explored. This ties in with the research objective of determining suitability of the technology for applicatory considerations. An overview of some of the challenges facing the technology such as GDPR are included and finally gaps within the research and the conceptual framework are included.

Clinical data integrity has been cited by the FDA and MHRA to be a concern of importance in the last number of years (Khin *et al.*, 2020), and has been identified as a historically and potentially ongoing issue (George and Buyse, 2015; Seife, 2015; Carlisle, 2017; Khin *et al.*, 2020). Issues that relate to data integrity include data transparency, traceability and accuracy. One area of concern relating to data integrity is data fraud. Data fraud is cited to be a rarely occurring phenomenon, however, George and Buyse, (2015) have suggested that clinical data fraud, while cited to be low, may in fact be much higher than is reported. This they argue can be due to a variety of reasons such as issues with definitional problems, problems surrounding study designs and under-reporting of fraudulent data in peer-reviewed papers. They consider the possibility that data fraud may in fact go undetected or unreported regularly (George and Buyse, 2015; George, 2016). Seife (2015) identified multiple clinical trials in a study of 57 trials that exhibited instances of data fraud including data falsification and false information submissions to regulatory authorities. Seife also states that although evidence of data misconduct such as fabrication is identified, it is reported to rarely be reflected in peer-reviewed literature (Seife, 2015). Apart from intentional data misconduct, data integrity suffers as a result of the sheer volume and complexity of the produced data in clinical trials which leads to problems relating to transparency and ultimately integrity.

Various studies and pilots have explored using BCT as a potential solution for improving issues of data integrity within clinical trials, but often lack any real observations or references to

instances of data fraud. Most of the studies which argue for improvements to data integrity utilizing BCT refer to the workings of George, (2016) and George and Buyse (2015) but often fail to explore beyond these studies. Furthermore, there is scarce literature examining the social perceptions and attitudes towards clinical data integrity concerns and BCT from an actual clinical data management perspective. In addition, much of the literature surrounding BCT proclaim that forcing stakeholders to use this technology would represent a challenge (Benchoufi and Ravaud, 2017; Wong, Bhattacharya and Butte, 2019), but do not elaborate on how these findings were arrived at or how they could be resolved. Most of the proponents for BCT implementation seem to originate from technology-based backgrounds rather than clinical settings. Whether BCT represents a "hype cycle" that fails to deliver innovative promises but instead stagnates in the "idea" phase remains to be seen (Mackey *et al.*, 2019).

The literature explores some of the strategies that have been carried out for implementing BCT into clinical trials and addresses some of the gaps found within the literature. These gaps encompass areas such as sociotechnical barriers of implementation and the challenges associated and includes brief mentions of GDPR challenges. As Blockchain is a new and nascent technology, most of the literature surrounding its implementation and benefits is exploratory in nature.

2.2 Bitcoin – An Introduction to Blockchain

In 2008, a person or persons under the assumed pseudonym Satoshi Nakamoto released a whitepaper: 'Bitcoin: A Peer to Peer Electronic Cash System'. This paper outlined a "purely peer-to-peer (P2P) version of electronic cash" that would facilitate online transactions between one party to another without the need for a third-party financial institution and "without relying on trust" (Nakamoto, 2009). Nakamoto identified several challenges and inherent weaknesses associated with third-party financial mediators, including the driving up of online transaction costs and fraudulent activities such as the double-spending problem (Nakamoto, 2009). In addition, central intermediaries represent a single-point-of-failure, meaning that if the intermediary was compromised the entire network could be at risk (Kuo, Kim and Ohno-Machado, 2017). Bitcoin is a digital currency also known as a cryptocurrency that uses cryptographic proof to carry out digital transactions directly between transacting parties, instead of placing trust in a third party (Vujicic, Jagodic and Ranđić, 2018).

Due to the anonymity of Bitcoin there was no way to verify and validate the transaction thereby preventing fraudulent activities. In order for Bitcoin to function as a decentralized public ledger without a central authority and to counteract the issue of fraudulent activities Satoshi Nakamoto conceptualized Blockchain, the technology which underpins the Bitcoin cryptocurrency platform (Vujicic, Jagodic and Ranđić, 2018). BC offers an alternative way of transferring data between parties without having to rely on a central mediator to validate the data transactions. The

anonymity of Bitcoin and Blockchain meant that the platform could be used to elicit illegal or nefarious online activities which has resulted in a negative perception of the platform by the public (Alshamsi and Andras, 2019).

2.2.1 What is Blockchain?

Blockchain is essentially a series of consecutively chained blocks which contain timestamped records of information that are maintained in a distributed public or private ledger across a network of computers also known as nodes. These nodes are linked together in a peer-to-peer (P2P) network; thereby negating the need for a trusted third party (Maslove *et al.*, 2018). Each participating node in the network contains a local copy of the ledger meaning that if one node or more disconnects, becomes corrupt or fails for any reason, this has no effect on the overall network (Sarmah, 2018). This removes the vulnerability associated with central authorities that represent a single-point-of-failure, as the database exists on multiple nodes spread across a network which occurs in real-time (Zheng *et al.*, 2017). This also means that if any local change is made to the BC it is replicated across the entire network resulting in each node updating their local ledger to match (Benchoufi and Ravaud, 2017; Vujicic, Jagodic and Ranđić, 2018). A Blockchain is considered immutable due to integrated cryptographic functions and the required validation of each transaction by each node (Sarmah, 2018). The differences between centralized and Blockchain networks are illustrated in figure 1.

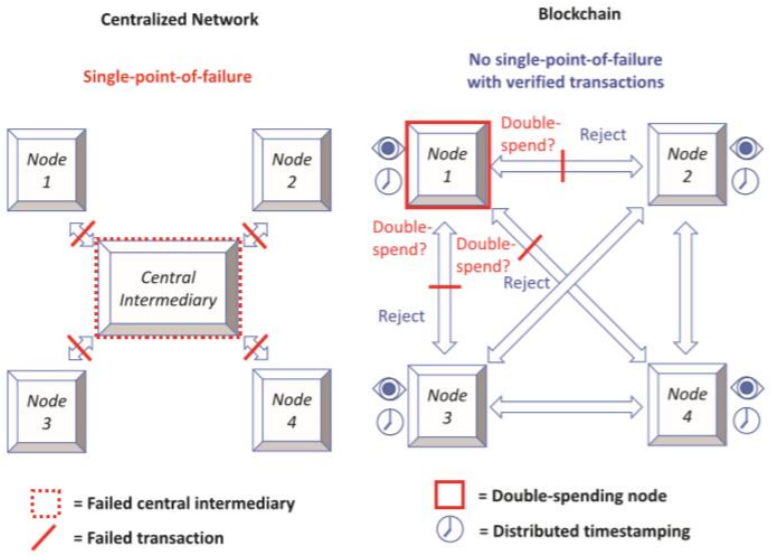

Figure 1: Centralized Network Vs Blockchain (Kuo, Kim and Ohno-Machado, 2017)

2.2.2 How Does Blockchain Work?

When a transaction is requested within a BC network it is uploaded to the P2P network of participating nodes. These nodes then validate the requested transaction and the user's status using mathematical algorithms and by reaching a collaborative consensus (Bashir, 2017). A transaction

in the case of BC can be a digital financial transaction or can be a transaction of information. A consensus is reached by various different methods depending upon the Blockchain architecture used. For example, within the Bitcoin BC, a consensus is reached once 51% of the nodes agree and validate the new block (Lin and Liao, 2017). Once the transaction is accepted and validated it creates a new block and adds it to the growing chain of blocks (Zheng *et al.*, 2017; Sarmah, 2018). This process is illustrated in figure 2.

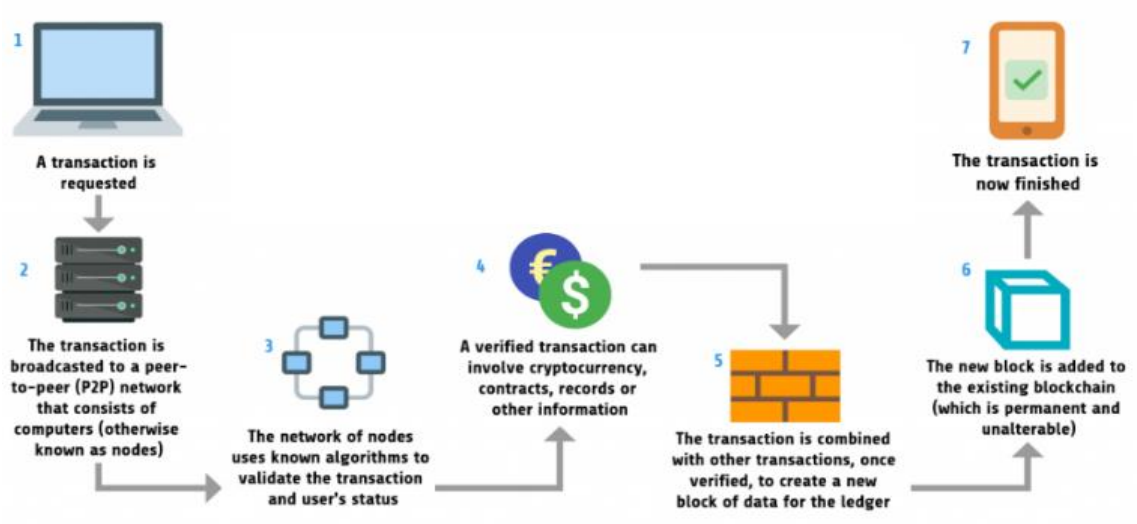

Figure 2: How BC works (IPSpecialist, 2015)

A block in a BC consists of data packages that contain a record of transactions or digital information. Each block is linked together by cryptographic hashes to form the chain of blocks (Hölbl *et al.*, 2018). A block will hold all transactions that happened within that designated timeframe. When a piece of digital information or transaction is created on a BC network it is stored within a new block. When a new block is created the information will go through a hashing function which will calculate and assign a hash to that block. A hash is a mathematically calculated unique code assigned to each block; essentially a digital fingerprint. The hash is a function that will convert the digital information contained in the block; letters and numbers, into an encrypted digital output at a fixed length through a mathematical algorithm (Zheng *et al.*, 2018). The hash codes function to connect the blocks in a specific order with each hash being unique to each block (Mann *et al.*, 2020). Each block is timestamped and also contains a reference (the hash) of the previous block in the chain (Hölbl *et al.*, 2018). When a new block is added to the BC, each node within the network will work collaboratively to come to a consensus on the authenticity of the block (Zheng *et al.*, 2017). Once a block is created it becomes a near immutable validated and encrypted set of digital information (Zheng *et al.*, 2017; Sarmah, 2018).

There are various types of Blockchains, for example, a public BC such as Bitcoins BC allows anyone to join the network and represent as a node. Private and permissioned BC's are managed and controlled by a single organization and require special permissions to join while consortium Blockchains are controlled by a group of organizations such as regulatory authorities, sponsors and contract research organizations. Private and consortium BC's are only partially decentralized but still possess all the benefits such as timestamping and immutability as a public BC (Maslove *et al.*, 2018; Mann *et al.*, 2020).

Figure 3 illustrates a typical example of a BC. Each block illustrated contains multiple transactional information's denoted as TX1, TX2 and TXn, the hash of the previous block, a timestamp denoting the time of creation of the block and associated information, and a nonce which a randomly generated number used to verify the hash (Zheng *et al.*, 2017).

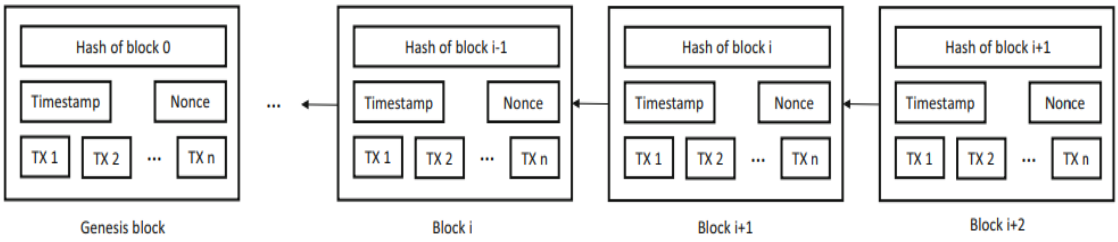

Figure 3: Typical example of a BC (Zheng et al., 2018)

2.2.3 Blockchain: A Tamper-Proof Ledger

Blockchain creates a tamper-proof ledger that is append only. If even a single character; letter or digit is tampered with in a block by a malicious user it will result in the change of the hash code of that block, meaning the subsequent hash of the following block would now be invalid. This is because the hash of the subsequent block in the chain is calculated using the hash of the previous block - which is now different. This results in a chain reaction of each block containing invalid hashes as each block would contain a hash that does not correspond with the hash of the previous block (Bashir, 2017; Zheng *et al.*, 2017). If a malicious user wanted to change the information contained in an already validated block, they would need to re-calculate the hash of the block. By design, this requires significant computing power and if a person were to successfully re-calculate the hash of the block, they would then need to recalculate the hash of the subsequent. By altering the hash of the original block, every hash for every block in sequence would need to be re-calculated, which is effectively impossible (Crosby, 2016).

Any identical documents will produce an identical outputted hash, however, if even a single character in one of the documents is changed, it will result in an entirely newly outputted hash. This means that if a person received the document, they can compare its hash with the hashed

document within the BC. The document within the BC is immutable, therefore if the hash of the document they received matches the hash of the document on the BC, they know none of the information contained has been altered in any way and the integrity of the document is intact (Mann *et al.*, 2020). The hash function in addition with the timestamp function ensures that information has not been interfered with and therefore provides a verifiable and tamper-proof historical record of all data transactions since the first genesis block (Wong, Bhattacharya and Butte, 2019). Due to its tamper-proof nature, BC can always ensure the integrity and sincerity of data by preventing data falsification after-the-fact (Benchoufi and Ravaud, 2017). This functionality of immutability has far reaching potential for the integrity of clinical data integrity which has historically been the subject of after-the-fact falsification and modification (George and Buyse, 2015; George, 2016).

2.3 Introduction to Clinical Trials

Clinical trials are complex processes that represent a key research tool in the advancement of medical knowledge and the improvement of patient care. They are fundamental to the discovery of new treatments for diseases while also providing the possibility for new medical interventions for the detection, diagnosis, and reduction of diseases such as cancer as well as infectious and autoimmune diseases (Pocock, 2013). Clinical trials test the safety and efficacy of a certain treatment such as a drug or therapy throughout a number of defined phases (Phases I-IV) with a defined number of voluntary participants for each phase (Shamley and Wright, 2017). They have a number of stakeholders who have a vested interest in the outcome of the trial, whether it is the sponsor who may invest large financial contributions in the hope of bringing a new potential drug or therapy to market or the patient themselves who may suffer from the condition the drug or therapy is designed to potentially treat (Shamley and Wright, 2017).

The key stakeholders involved in a typical clinical trial consist of the clinical team, sponsor, IRB/IEC ethics committee, regulatory authority, study participants and sometimes a contract research organization (CRO). The study participant is arguably the most important stakeholder in a clinical trial and are voluntary eligible candidates who will sign an informed consent to participate in the study. The ethics board are responsible for ensuring that the trial remains ethical and in accordance with GCP guidelines while protecting the welfare and human rights of the participants. The sponsor may be a pharmaceutical company or institution that initiates and manages the clinical trial and may sometimes contract a CRO to manage or perform certain aspects of the study. The relevant regulatory authority is responsible for ensuring clinical trials meet quality, safety and efficacy requirements and overall are responsible for the approval of a new drug (Shamley and Wright, 2017). The Clinical Data Manager is responsible for the overall clinical data management process (Krishnankutty *et al.*, 2012). All of these stakeholders must collaborate on a range of trial activities such as data management monitoring, regulatory

reporting, site monitoring, safety reporting and medical writing amongst others despite being geographically dispersed (Shamley and Wright, 2017; Triall, 2019). The various stakeholders are illustrated in figure 4 (Benchoufi and Ravaud, 2017). Clinical trials are separated into phases which represent separate trials with each phase intending to answer specific research questions. Each phase will have progressively more participants and time taken to completion. Phase III can involve thousands of participants over several years (Shamley and Wright, 2017).

In 2010 there were 82,865 registered studies across the United States and in a further 213 countries. In contrast, as of May 14, 2020, there are over 325,000 registered studies (ClinicalTrials.Gov, 2020). A study carried out by Wong, Siah and Lo (2019) found that almost 14% of all drugs within clinical research trials eventually gain FDA approval but this still means 86% of drugs fail to win market approval. They also found that non-oncology trials can run up to 7.2 years while oncology trials can run up to 13.1 years (Wong, Siah and Lo, 2019). As the number of registered studies increases so too does the complexity of the trials, the number of participants and the monetary investments required by sponsors. A typical full clinical trial can cost in the range of $10 to $30 million and can go even higher in some cases, while the current estimated cost for bringing a new drug or therapy to market is between approximately $1 to $2 billion. It has been shown that cost of bringing a new drug to market rises by up 9% each year (Triall, 2019). Two of the primary cost drivers associated with clinical trials are data management and the ongoing requirement of compliancy with increasing regulations (Sertkaya *et al.*, 2016).

2.3.1 Data Management in Clinical Trials

Clinical trials generate significant amounts of scientifically and medically relevant data regardless of the outcome of the trial, and therefore play an instrumental role in the advancement of medicine and patient care (Pocock, 2013). Clinical data is used by regulatory authorities for the approval decision making of new drugs and treatments while additional stakeholders such as research institutions, medical devices, pharmaceutical companies and prospective patients alike all rely on clinical data being accurate, transparent and traceable (Maslove *et al.*, 2018). Regulatory authorities rely on the reproducibility of clinical data in order to reconstruct the trial to confirm the integrity of the data (Khin *et al.*, 2020). However, accuracy, transparency, traceability and reproducibility of clinical data has historically been a fundamental and recurring issue within clinical research as approximately 80% of clinical trials are cited as being non-reproducible (George, 2016; Benchoufi and Ravaud, 2017; Khin *et al.*, 2020). The non-reproducibility of trials is due to a variety of data-related problems such as erroneous data, data fraud and misconduct and data misrepresentation (Benchoufi and Ravaud, 2017; Wong, Bhattacharya and Butte, 2019) As the complexity of data increases alongside with the number of active trials, it is evidently becoming increasingly important to secure the accuracy, transparency and traceability of all clinical data that is produced throughout a study.

A typical clinical trial consists of numerous concurrently running data streams that are intertwined to produce complex and substantial quantities of data that require constant monitoring and analysis by multiple involved clinical staff. This includes data involving recruitment and retention, consent forms, data monitoring and analysis, security reporting and protocol management (Benchoufi and Ravaud, 2017). Figure 4 illustrates an overview of the various data streams and involved stakeholders associated with a typical clinical trial. When managing large amounts of clinical data that is subject to change it is essential to ensure that there is an audit trail that is immutable and historically accurate, however, this currently represents a major issue for regulatory authorities within clinical trials (Hume *et al.*, 2017; Khin *et al.*, 2020).

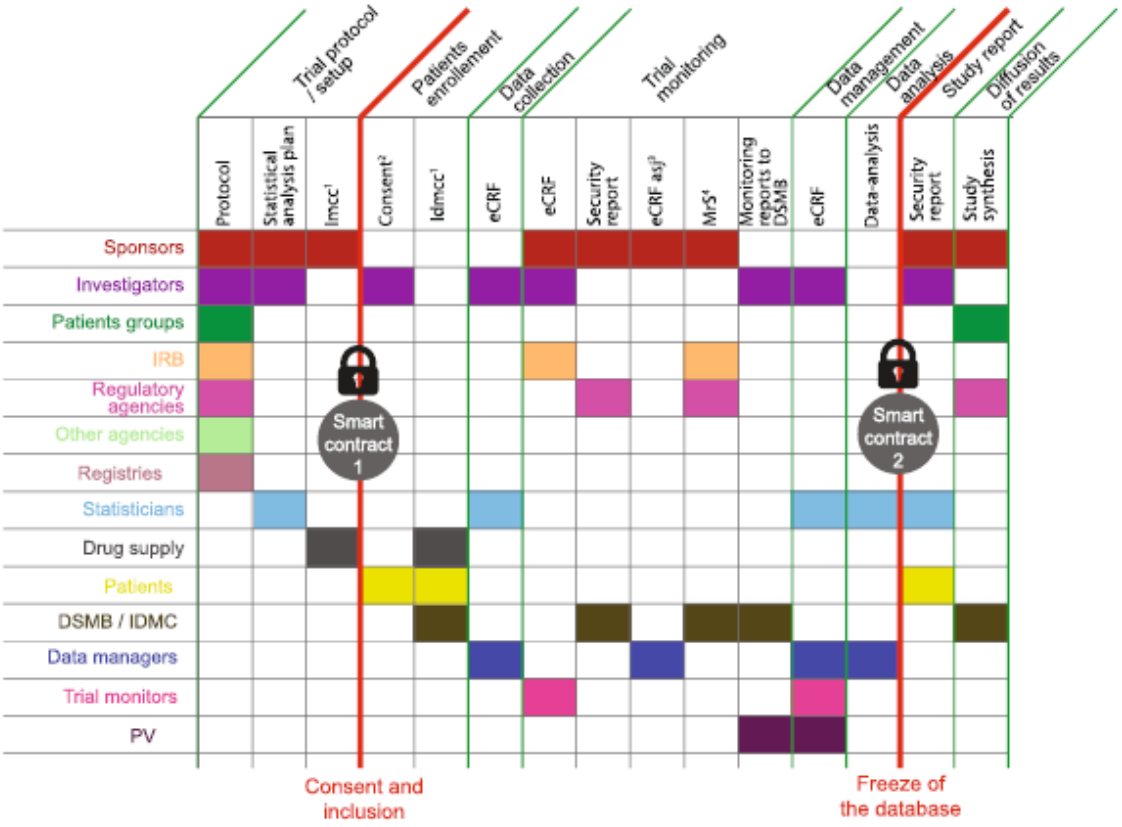

Figure 4: Overview of clinical trial data streams and relevant stakeholders (Benchoufi and Ravaud, 2017)

2.3.2 Clinical Data Integrity Concerns

Maintaining clinical data integrity is intrinsically important in clinical research as erroneous data could result in a new drug being rejected for market approval by the relevant regulatory oversight bodies. It could also result in additional studies being carried out and can result in adverse reputations for pharmaceutical companies (Khin *et al.*, 2020). Inconsistent, unreliable and inaccurate data can incur monetary cost implications for sponsors and may result in aspects of the

trial requiring repeating thereby increasing the length of the trial (Gupta, 2013). Trust in the integrity of clinical data is vitally important not just for the outcome of a trial, but also for post-trial completion and subsequent regulatory approval of a treatment and subsequent post release to market. A considerable amount of sensitive data must be monitored and validated while shared and should be made accessible in a reproducible and transparent manner. In addition, it must be guaranteed that data remains private and inviolable (Albanese *et al.*, 2020). The International Council for the Harmonization of Technical Requirements for Pharmaceuticals for Human Use - Good Clinical Practice (ICH-GCP) offer guidelines on how clinical trials data should be managed, monitored and carried out and state that "All clinical trial information should be recorded, handled, and stored in a way that allows its accurate reporting, interpretation and verification" (ICH, 2018). Despite GCP guidelines, clinical research has been the subject of numerous issues surrounding how data is managed, stored and shared (Seife, 2015; George, 2016; Benchoufi and Ravaud, 2017). As clinical trials become larger and more complex, they produce larger volumes of data that require labor-intensive management on an on-going basis, and this subjects the data to the possibility of data errors, whether unintentional or malicious (Wong, Bhattacharya and Butte, 2019). Data errors and inconsistencies can be the result of honest erroneous data entries or can be the result of data misconduct such as fabrication and falsification (George and Buyse, 2015; George, 2016).

2.3.3 Clinical Research Data Fraud and Misconduct

Clinical research data fraud and misconduct is not an unheard-of phenomenon within clinical research. George (2016) suggests that researchers may commit data fraud or data misconduct for personal and/or financial gain or can be the result of the mounting pressure placed on researchers to meet trial objectives under increasing regulatory requirements and financial constraints (George, 2016). Researchers as well as sponsors have strong incentives to publish positive data results which can have a damaging effect on public health if the data is unreliable and erroneous (Scarso Borioli and Couturier, 2018). A study carried out by Seife in 2015, found that out of fifty-seven published clinical trials included in the study that were subjects of FDA inspections, 22 trials were found to have a falsification or submission of false results, while 42 trials were found to have inadequate or inaccurate data recordkeeping (Seife, 2015). Seife reviewed several published inspection documents by the FDA and identified "major violations of GCP, including allegations of fabrication and other forms of misconduct" (Seife, 2015). Additionally, according to Seife, departures from GCP are often not disclosed by the FDA and findings are rarely cited or reflected in peer-reviewed journals even if evidence exists to support misconduct or erroneous data.

George (2015) illustrates several high publicity cases in which research investigators committed clinical data fraud. One high publicity case included a researcher who was found to have

committed data fraud in 172 research papers which included 126 papers that contained fabricated reporting of results (George and Buyse, 2015; Carlisle, 2017). The literature is replete with historical high profile cases of clinical data misconduct and fraud (Gupta, 2013; George and Buyse, 2015; Seife, 2015; George, 2016). Furthermore, Carlisle (2017) identified possible data misconduct in 90 trials when reviewing data from 5,087 published clinical trials by incorporating statistical tools to identify data inconsistencies. He noted that the statistical data patterns recorded in the 90 trials were highly unlikely to occur by chance and indicated nefarious instances of data manipulation and fabrication (Carlisle, 2017). It is clear from the literature that data misconduct is not an uncommon occurrence and the likelihood of incidences may in fact be higher than expected or reported.

However, despite these findings, actual prevalent proof of data misconduct is in fact cited to be low as the true rate of occurrences can be problematic to estimate. This is supported by George (2016) who describes some of the possible reasons behind this. For example, determining prevalence and proof of misconduct can be challenging as it can be difficult to clarify who the study population is when investigating misconduct. Numerous surveys and investigations have attempted to determine prevalence in recent years resulting in disparate and conflicting conclusions (George and Buyse, 2015). The data produced from the study carried out by Seife (2015) was not quantitative but descriptive. This was due to the limitation of the procured data coming from what was publicly made available by the FDA, which was often redacted, and which also was updated infrequently on the FDA website (Seife, 2015). Clinical research data misconduct can vary in severity and if the regulatory authority such as the FDA discovered evidence of a violation, it does not mean that it is required to be publicized or included in an article and often FDA documents are redacted making it difficult to determine which trials were found to have data misconduct (Seife, 2015). It must also be considered that anomalies of data found within literature could in fact be attributed by misinterpretation or simple human input data errors, or honest transcription mistakes. Nevertheless, it may difficult to prove widespread prevalence, but its occurrence is fact.

Although it may be difficult to measure and quantify the extent of actual clinical data misconduct, the probability of it occurring within a high-pressure environment is not outside the realm of possibility. It can be surmised that possible minor data instances of misconduct may go undetected on a regular basis. In addition, clinical data misconduct may not always be the result of an intentional action, but rather the result of ineffective and poor management (Gupta, 2013). Poor data management systems or high-pressure environments could result in carelessness of data entries by clinical staff. Proof of data misconduct in clinical trials may be low, but honest data errors, honest mistakes, and poor data management is reported to be a widespread issue (George and Buyse, 2015; Benchoufi and Ravaud, 2017). Improving data integrity is an ongoing endeavor

as regulatory bodies such as the FDA, WHO, the European Medicines Agency (EMA) and the MHRA who have all released guidance documents pertaining to the necessity for the improvement of clinical data integrity in the last number of years (Khin *et al.*, 2020). Regardless of the original intent of data misconduct, whether malicious or accidental, clinical data integrity is ripe for an effective, transparent and immutable data audit trail overhaul. Erroneous data results in financial and reputational damage which can impede clinical trials objectives and any attempt to improve its integrity should be explored.

2.4 Blockchain-Enabled Clinical Trials

Due to Blockchains immutability, it can in theory greatly reduce the possibility of data misconduct occurring as any data inputs can be traced back to the origin and the data cannot be altered once uploaded due to the cryptographic hashing functions. Real-time data access in a BC-enabled clinical trial will allow relevant stakeholders to monitor data continuously and this in theory would have significant impacts of the quality of produced data. With current clinical data management systems, it is becoming increasingly difficult to trace data back to its origin or to see who has also accessed it or modified it a point in time. This is due to data being stored across multiple locations and in a centralized manner. With clinical data stored on a BC, data provenance is greatly enhanced as data can be traced back to its origin with full transparency and trust. It has however, been suggested that Blockchain is facing an innovation plateau caught in a 'hype-cycle' as is the case with many new technologies (Mackey *et al*., 2019). However, BC has a proven track record in terms of its applicability with cryptocurrency trading which by design is 'trust-less'. One can be absolutely assured that data on the BC is 100% transparent and accurate and if this proven track record can seep into clinical data management it could potentially overhaul how all clinical data is stored and shared going forwards.

2.4.1 Ensuring Data Integrity through Blockchain Regulatory Auditing

Regulatory authorities such as the FDA are responsible for ensuring that clinical trial methodologies are conforming and adhering to GCP guidelines and facilitate this by auditing on-site records and data (Hirano *et al.*, 2020). A lack of transparency and traceability of data increases the likelihood of erroneous data going undetected and the FDA stated traceability and transparency to be among their top predominant data issues that need to be addressed (Hume *et al.*, 2017). Sponsors must be able to prove data provenance and integrity and be able to address queries from oversight bodies in order to maintain integrity of data throughout the trial. This is a time consuming, expensive and difficult process for both the sponsor and the regulatory authority (Mackey *et al.*, 2019). Audit trails provide the regulator with the necessary data to ensure the trials data integrity. Khin *et al.*, (2020) whom reviewed the discussions between the FDA and MHRA outline that audit trails are critically important for capturing data collected from a study and should clearly outline who added data, when it was added and what was added. Furthermore,

appended data should not obscure the original data and data should not be deleted once entered. They also observe that audit trails should be accessed and reviewed throughout the study in an easily accessible manner by the regulator to ensure ongoing data integrity (Khin *et al.*, 2020). Blockchain possesses the architecture that can facilitate this and can transparently provide full data provenance from origin to completion to concerned regulatory authorities on a real-time basis. BC could enhance the trust in data allowing regulators to streamline the approval process.

The traditional practice in clinical research for ensuring data integrity and accuracy has been 100% source data verification (SDV). SDV is when data collected in a case report form (CRF) or other data collection systems are compared with the original source of data or information to confirm the accuracy of the transcription of the information. SDV has been estimated to cost up to 25% of the clinical trials entire budget. However, as clinical trials grow in size and complexity it is becoming increasingly difficult to apply 100% SDV. One of the reasons for this is the increasing regulatory requirements imposed on clinical research (Hirano *et al.*, 2020). Regulatory bodies such as the FDA released a paper suggesting the use of alternative methods to 100% SDV with an emphasis on the consideration of electronic monitoring approaches (FDA, 2013). Wong, Bhattacharya and Butte, (2019) recognized the need for increased clinical data integrity and correlated this with the need for a real-time auditing methodology for regulators. They modelled a private BC based on a simulated prototype phase II clinical trial which incorporated real data from a previously completed trial. Control of their BC was given to the regulator (hypothetically) who in theory would carry out real-time auditing of data from the data uploaded to the BC. All data created since the genesis block would be available in an immutable and chronological manner and with BC's function of cryptographic hashing the integrity of the BC's data can always be maintained. All additions to the append only BC can be traced back to the author via digital signatures and hashing. This enables true transparency and traceability of data (Hume *et al.*, 2017). There are several limitations to this proof-of-concept study. For example, forcing all stakeholders or participating parties can be a challenge due to sociotechnical barriers. Only data recorded on the BC could be considered for regulatory purposes, this would not include any offline transactions and entering data on a BC does not eliminate data fabrication or falsification at the point of origin (Wong, Bhattacharya and Butte, 2019).

Maslove *et al.*, (2018) have also suggested the use of a private BC to streamline regulatory auditing with their POC called BlockTrial. Some of the limitations outlined in the study also included sociotechnical barriers for stakeholders in adopting a BC enabled clinical trial data infrastructure. Hirano *et al.*, (2020) also identified data integrity as being a key issue and examined the impact of incorporating BC to enable real-time auditing by oversight bodies. From these studies it is evident that BC could streamline and even accelerate the drug approval process

while providing regulators with transparent data that they can query in real-time which could disincentivize data misconduct or even careless data erroneous inputs.

2.4.2 GDPR & Blockchain Tensions

As with any nascent technology, challenges are numerous concerning its adoption and implementation. One predominant issue facing BCT is the European General Data Protection Regulation (GDPR). GDPR is based upon the assumption that there is a 'data controller' for each personal data point. This contradicts BC's platform in that there is no single 'data controller' but instead there exists multiple controllers. This raises the concern of ownership and accountability of clinical data stored on a BC (European Parliament, 2020). This may be bypassed in CT research if the BC was private and permissioned and central control was given to one stakeholder such as the sponsor (Wong, Bhattacharya and Butte, 2019). A second concern is that data can be modified or erased under GDPR regulations in order to comply with any legal requirements (European Parliament, 2020). However, BC is built upon its immutability an append only architecture thereby modification or erasure of data is not possible (Zheng *et al.*, 2017). Data contained on a blockchain is usually encrypted or hashed (particularly in permission-less BC's) and there is still ongoing debate as to whether encrypted data qualifies as personal data. The tensions between BC and GDPR are extensive and intricate, particularly when personal health is concerned. In addition, there exists not one 'Blockchain' but many types with varying and disparate features and architecture (European Parliament, 2020). Therefore, it may be the case that there is only one standardized BC that is simultaneously designed to handle all areas of CT research from consent to data sharing while also remaining compliant with data protection laws. However, studies in the area of CT data remain underdeveloped and it may transpire that there may be a need for several BC's to handle each individual area of CT research rather than one single BC. In conclusion it is evident that there is much research to be carried out regarding BC and data protection laws and at present this will most likely represent a key challenge.

2.5 Key Findings

Below is a list of the key findings discovered from the literature:

- There multiple clinical data issues identified from the literature. These involve data issues relating to traceability, transparency, quality and auditability which all contribute towards the general integrity of clinical data
- There have been multiple recorded incidences of data fraud and data misconduct contributing towards the lack of general trust in the validity and reliability of clinical trial data
- There is a regulatory recognition that clinical data integrity is an ongoing concern that needs to be addressed

- Blockchain has been considered for the application of all areas of data management within clinical research from consent approval and management to recruitment of potential candidates to the management and traceability of clinical data throughout the trial and post completion of the trial
- There may be negative perceptions to the technology from the public which may weaken its brand and may act as an obstacle for its successful integration within clinical data management
- There are currently no clinical trials that have implemented the technology indicating the technology is still in its infancy of consideration for implementation and there are most likely significant challenges that must be overcome, however, these are seldom mentioned throughout the literature
- Blockchain is complex and intricate and this may act as a inhibitory factor for its uptake by clinical staff
- Due to the emerging and nascent nature of this technology, it can be presumed that there is a lack of BC experts who are also proficient in regulatory clinical data compliancy laws
- BC is potentially caught in a 'hype-cycle' and may never come to fruition

2.6 Gaps in the Research

There are extensive gaps in the literature pertaining to the challenges that face BC's implementation with clinical data management. Although it is not the intention of this study to examine the challenges facing BCT, these will be explored lightly in the primary research, as this could form the basis for future work and represents a key area which should be explored. There also exists gaps in the literature concerning the uptake of this technology amongst professionals within the clinical research industry. It may surface that stakeholders are simply not comfortable using a technology that has a negative association with something like Bitcoin, or they may simply not trust the "trust-less" technology for the management of personal health data. This general mistrust may come down to a lack of understanding of how the technology works and this will form the basis of the research.

2.7 Conclusion of Literature

It is clear from the extensive studies and literature that BC holds far reaching potential benefits for the traceability, auditability, integrity, security and transparency of clinical data. Various POC studies have proposed BC for almost all aspects of clinical research, however one area that is scarcely examined is the attitudes and social technical adoption of BCT. There may exist social barriers to its adoption as many researchers may not even be aware of this potential "disruptive" technology. Most of the literature which professes the potential benefits of BCT for the

improvement of clinical research data integrity refer back to the studies carried out by George (2016) and George and Buyse (2015), however, often seldom explore beyond these studies. There is no literature or research that is inclusive of actual clinical data managers or opinions of clinical data and integrity and this presents a gap that should be explored.

While it is clear there is a market for BCT in clinical trials, there still exist no real-world applications within clinical studies at the time of this research and no research to evaluate the attitudes and receptiveness of the technology by the clinical research industry. The application of BCT may be frivolous if its uptake and implementation is rejected by researchers and stakeholders alike. In addition, data integrity encapsulates a wide range of sub-topics ranging from data fraud to honest data errors and poor data management and BCT may not be the solution for these issues. Deeper real-world investigations are required to determine and examine the attitudes and perception of this technology in the effort to improve clinical data integrity. Blockchains association with cryptocurrency may potentially yield an unfavorable reputation as untrustworthy, as cryptocurrencies are often deemed intrinsic in eliciting illegal online transactions where users remain anonymous (Bashir, 2017). BC is considered "trustless" due to its cryptographic hashing functions, but ironically the social perception of trust for the technology may remain an obstacle for its implementation. There may be a distinct lack of understanding of the true potential of BC and its potential for enhancing and streamlining the management of clinical data.

It is important to note that the aim of this study is not to identify or prove instances of possible research misconduct or to analyze the possible potential repercussions and consequences of it, but rather to highlight the possibility of occurrences, and to investigate how a technology such as BC could be implemented to potentially prevent the possibility of it occurring in the future. It is also not the intention of this study to highlight possible investigational inconsistencies by regulatory authorities into clinical misconduct, but instead to outline a new and untested method of data auditability that could be implemented to improve data integrity and real-time auditing. Pharmaceutical companies are faced with an increasingly difficult regulatory landscape in which to deliver trials designed around patient safety in accordance with GCP guidelines (Khin *et al.*, 2020). There is pressure from sponsors to recruit and enroll target study participants as well as to successfully reach trial objectives and timelines and all within budgetary allocations. There is pressure from research ethics committees for trials to align with ethical and moral standards and even pressure from patients who are placing their personal dependencies upon the successful outcome of a trial (Shamley and Wright, 2017). With such a high number of variables it is reasonable to surmise that there exists various cases of questionable research practices and misconduct as well as a range of honest mistakes. However, improvements in clinical data management represents a key area in which many solutions for data errors, intentional or

malicious, could be derived from and Blockchain could potentially be the solution for some of these issues.

2.8 Conceptual Framework

One of the key findings from the literature is the lack of research into the social attitudes of clinical researchers to BCT and clinical data integrity. This will form the fundamental basis for the research undertaken. BCT proponents profess the benefits that BC holds for the challenges associated with clinical research data integrity, yet there exist no real-world applications at present. Authors of BC advocacy in alleviating data integrity in clinical trials tend to loosely base their frameworks and ideologies on the studies carried out by George (2016) and George and Buyse, (2015) but fail to explore whether BCT can provide the solution for the proposed ongoing problematic data integrity trends. To date, from the examined literature, there has been no qualitative and quantitative research from directly involved clinical staff such as Clinical Data Managers that evaluate and examine the need and or requirement for such a technology.

Chapter 3

Research Methodology

3.1 Topic Outline

The objectives of this research are to gain a deeper understanding to some of the predominant issues surrounding clinical data integrity, to evaluate the current knowledge, awareness and perceptions of BC by clinical data staff, and to explore the suitability of BC in addressing the issues pertaining to clinical data integrity. Through a mixed method approach the study aimed to explore first-hand from clinical data research staff what they believe some of the primary concerns are that are associated with data integrity and how they feel they could be improved. This ties in with the research objectives of firstly examining the perceived issues relating to clinical data integrity. Perceptions and awareness of BCT will be evaluated and following this, the application of BCT will be explored in relation to clinical data integrity.

This chapter outlines the chosen research philosophy, strategy, purpose, process and research logic. Also included are the research methodology, sources of data, the sampling methods, analysis of data, and an outline of any access and ethical issues. A pragmatist inductive approach was taken with this study of which the research purpose was considered exploratory. The research method used was a complementary quantitative and qualitative mixed methods approach. The sources of data consisted of an online survey while the sampling methods employed were non-probability and purposive homogenous sampling. A thematic analysis approach was then taken in order to analyze the data produced from the qualitative questions and the quantitative data was analyzed in conjunction with the qualitative data.

3.2 Research Philosophy

The term research philosophy refers to a system of beliefs and assumptions about the development of knowledge and relates to how a researcher's intrinsic perceptions and beliefs influence their interpretations and understandings about the nature of the world. This ideology of beliefs and assumptions play a role in the way in which researchers gather, analyze, and use data, as a researcher strives to develop knowledge in the particular field in which they are studying (Saunders, Lewis and Thornhill, 2015). Throughout research, the researcher formulates a series of types of assumptions either consciously or subconsciously (Burrell and Morgan, 1979). These include:

- **Ontological** assumptions about the nature of reality
- **Epistemological** assumptions about human knowledge
- **Axiological** assumptions about human value and ethics

The assumptions made by the researcher will overall effect and shape the researchers understandings of the research questions, methods used, and the subsequent interpretation of the findings (Saunders, Lewis and Thornhill, 2015). Affecting these assumptions is a continua which are two extremes known as objectivism and subjectivism. Objectivism argues that social reality

is external to us and that social entities exist independently within one general reality. In objectivism the researcher is typically detached from the research and employs quantitative research methods, implying that facts and figures garner the best scientific evidence. Employing subjectivism means a researcher is typically attached to the research carrying out qualitative research (Saunders, Lewis and Thornhill, 2015). Subjectivism argues that social reality is formulated from the perception and consequence of people and that there are multiple perceived and experienced realities. Subjectivism involves the researcher to be more inclined to be interested in the various opinions and narratives to account for different social realities (Saunders, Lewis and Thornhill, 2015). Taking a subjective approach allows the researcher to be introspective of their own values and while they cannot detach from their values, they can incorporate them into their research. The subjective researcher will ultimately endeavor to understand the differing realities in order to make sense of the situation and to instill a sense of meaning rather than empirical evidence such as that of the objective researcher (Saunders, Lewis and Thornhill, 2015).

There are four major types of research philosophies (Saunders, Lewis and Thornhill, 2015).

- **Positivism** can be described as taking the natural scientist philosophical that entails working with a social reality that is observable producing law-like generalizations. Positivists focus more on collecting empirical evidence and facts rather than socially attributed meanings and impressions. Positivists will generally lead a quantitative study to test an existent hypothesis or develop one with existent theory in objective manner.
- **Realism** relates to scientific enquiry and describes the truth of reality being what our senses show us. Realism refers to objects existing in reality independently of the mind's perceptions and beliefs, in other words, their conceptual scheme. There are two subtypes of realism; direct and critical realism.
- **Interpretivism** can be described as a form of positivism but with a subjective perspective and advocates the necessity to understand the differences between different social actors and emphasizing the way in which each human creates and experiences different social realities. Humans act out their roles according to the meaning they attribute to these roles. Interpretivism researchers will adopt qualitative methods.
- **Pragmatism** researchers reconcile objectivism and subjectivism and often combine both positivism and interpretivism within the scope of their research. A pragmatist philosophy argues that the research question is the most important determinant of the research philosophy and will use whatever methods and philosophies necessary to further their research. They believe there are multiple ways to interpret reality, that multiple realities may exist and that multiple methods may be used when undertaking research, and that

that there is no single philosophy or set of ideologies that can best describe the entire picture.

For the purpose of this research a pragmatist research philosophy was chosen as the research paradigm underpinned by epistemological assumptions. The dichotomy presented between subjectivism and objectivism is not recognized by the pragmatist researcher, instead the researcher seeks to reconciliate the two contrasting concepts (Saunders, Lewis and Thornhill, 2015). A pragmatist approaches aligns particularly well with the mixed methods approach adopted as part of this research as the pragmatist researcher uses whatever means necessary to garner the necessary information and data that best facilitates the research question. Each participant within the study will have their own set of ideas and beliefs pertaining to the issue of concern and each may have alternative practical methods or solutions to address these issues. This approach allowed the researcher to become immersed in the perspective, interpretation and experiences of the clinical research staff which aid the researcher in further developing their knowledge of the issues and concerns without being confined by the contentious ideologies of truth and reality. Epistemologically, participants have their own experiences and knowledge that govern their responses to questions pertaining to clinical data integrity and BC. A pragmatist approach allowed for the collection of these experience and knowledge-based answers to form the basis for the pragmatic need for adopting BC to meet the data integrity issues outlined.

3.2.1 Research Design

The overall research approach for this dissertation was considered exploratory. This approach was adopted in order to gain a deeper insight into clinical data integrity from experts working directly within the industry – an area which has been seldom explored in relation to BCT. The exploratory research approach includes a search and review of the literature pertaining to clinical data integrity issues and BCT's potential applicability and relevance within clinical trials. This was followed by carrying out an online survey with clinical staff directly involved in clinical data management such as Clinical Data Managers to gain a deeper insight and perspective. Exploratory studies are useful in clarifying an understanding of a phenomenon or problem (Saunders, Lewis and Thornhill, 2015). In this case, the application of BCT in addressing clinical data integrity concerns from clinical staff will most likely not provide conclusive evidence that this technology can actually improve data integrity. But it was important to explore the necessity of introducing an entirely new data management system when managing and storing such sensitive clinical data. The insights gained from clinical staff will be useful for future considerations of implementing this technology in a real-world case. In addition, the social acceptances and perceptions of introducing new technologies such as BCT within clinical workplaces to handle data will also be beneficial for future related studies. A personal investigation is adopted which consists of an online survey that uses both quantitative and qualitative style questioning.

3.2.2 Research Process and Research Logic

There are two main ways in which data can be collected, these are qualitative and quantitative. Quantitative methods are often aligned with the positivism paradigm. Quantitative methods are often employed to directly examine relationships between variables utilizing standardized empirical and statistical recording methods of data. Qualitative methods are often aligned with the interpretive paradigm as researchers employ these methods to gain a deeper understanding and clarity of the subjective and socially constructed meanings regarding the issue or phenomenon being studied (Saunders, Lewis and Thornhill, 2015). Qualitative data collection methods are non-standardized, unlike quantitative methods. In addition, research may be inductive, deductive or abductive. Inductive research is associated with developing theories as a consequence of the analysis of the discovered data and aims at gaining insights into the meanings attached to situations or events. Induction research is often aligned with the interpretivist paradigm. Deductive research involves the development of theories and hypotheses and subsequently testing them by collecting appropriate data to confirm or disprove the hypothesis (Saunders, Lewis and Thornhill, 2015).

For the purpose of this research a mixed methods inductive research process was adopted by using an online survey. By employing a mixed methods approach, the pragmatist researcher can add a level of complexity and depth to their research by incorporating both quantitative and qualitative mixed methods (Saunders, Lewis and Thornhill, 2015). A qualitative approach was selected in order to gain a deeper insight into the proposed issues of clinical data integrity. This approach unearths some of the primary concerns relating to data integrity and highlights some of the attitudes and perceptions of clinical staff in adopting BCT to address some of these concerns. Due to the limited participants involved in the study, the qualitative approach is supported by a simultaneous complementary quantitative approach. The reasoning for this choice was to preemptively tackle the issue in which a participant had limited or no prior knowledge of BCT. For this reason, some of the questions were multiple-choice, hence a quantitative approach. This allowed the participant to choose an option they deemed to be the most satisfactory answer based on their own professional experience and knowledge while albeit possibly possessing limited knowledge of BCT. Quantitative and deductive research are often associated together within research (Saunders, Lewis and Thornhill, 2015). However, although quantitative questions were incorporated into the survey, the researcher did not consider these questions to be deductive as they functioned to support and complement the qualitative data. It was not the intention of the research to implicitly lay out empirical data that represented a target population, but to inductively apply both quantitative and qualitative questioning styles to complement the responses from both data sets to support the discoveries from the literature.

The multiple-choice options were derived from discoveries that were found within the literature review and form the most relevant subject areas of applicability. Each multiple-choice question had an 'other' option. If the participant felt that none of the provided options best correlated with the question and their own personal experiences, they could then add their own personal response under the 'other' option. This strengthened the mixed method approach while both facilitating those who have experience and prior knowledge of Blockchain with those who do not. Furthermore, limitations of incorporating quantitative questions in an online survey setting were recognised, especially in the case of participants having a lack of prior knowledge of BC as this could lead to uninformed responses which could affect the reliability of the survey. Part of this research was to determine awareness of this technology and participants were chosen who were most likely to be have been aware of BCT based upon their professional background. However, during the research process where a participant has not heard of BCT to any extent, a brief unbiased explanation of BCT was provided to participants. The ideal scenario was to interview participants with extensive knowledge and insight into both BCT and clinical data integrity, however from secondary research it was foreseen that this may not be the case as BC is still such a new technology and relevant participants were not readily available at the time of the research to engage in the study. Therefore, based on these findings, the survey questions were designed with both quantitative and qualitative methods in mind so both approaches could support and overcome any weaknesses that may be found by using just one approach.

Due to the inductive and exploratory approach no theories or hypotheses were applied at the start of the research. Using an inductive approach will allow a researcher to identify emerging themes throughout progression of the research by recognizing socially contextual meaning and understanding the perceptions study participants have relating to the research questions and objectives (Saunders, Lewis and Thornhill, 2015). Taking this approach can prove problematic as inductive research can often take significant time to uncover and identify trends on which to build upon, however, based on the researchers understanding of the study, this was the most appropriate approach to take in order to analyze the data. Although surveys are often associated with a deductive nature, the survey employed as part of this research is designed to be inductive in nature.

3.3 Sources of Data

The collection of data was gained from an online issued survey with clinical data research staff members directly involved in clinical data management and data inputting. The survey was designed to evaluate how clinical data suffers from issues of integrity and the main causative factors that contribute to integrity problems. Surveys can be useful mixed methods approach of gaining qualitative and quantitative data as they can encompass both open-ended and closed questions (Saunders, Lewis and Thornhill, 2015). Due to the limited number of respondents, this survey approach cannot be said to be reflective of the population of interest, however, this is not the intention as the questions were designed to be inductive by nature. Open-ended questions allow participants to share their own personal answers based on their own perceptions, experiences and personal knowledge (Saunders, Lewis and Thornhill, 2015). This lends towards the exploratory nature of this research. Quantitative questions included in the survey were designed to be complementary with the qualitative questions and vice-versa. The survey was shared with those for whom the research was aimed towards namely those currently or have been in the past involved with clinical data management.

As there are no real-world applications of BCT in use in a clinical trial setting identified at the time of this research, questions regarding BCT include a brief introduction to the key points and potential applications of the technology. Awareness of the technology may be low, and this was accounted for when designing questions relating to its functionality and implementation. Furthermore, while on the surface, proponents of BCT profess that BCT holds enormous 'obvious' benefits for clinical trial data integrity, this may not be the real-world case. While the researcher is slightly biased in favor of this; based on the findings from the literature review, the questions regarding BCT are structured as best as possible as not to appear to be biased in favor of its implementation. Questions regarding BCT predominantly seek to outline the attitudes and awareness of its potential application and this links with overall research objectives of determining the perception of the technology in relation to CT data integrity concerns.

3.4 Sampling Methods

This was a mixed methods exploratory research with no sample frame and a non-probability sampling technique was used as the sampling method. Homogenous purposive sampling was used to select cases that would best represent those who could meet the research objectives. The researcher's judgment was used to purposively select participants using clear criteria of those of whose profession involved clinical data management and analysis as these participants best represented the sample population who could provide meaningful data regarding both clinical data integrity and data relating to BC. Potential participants were selected based upon their professions by searching on social media platforms such as LinkedIn and Facebook. In addition, online groups such as Clinical Data Management groups were identified and from there further

participants were purposively selected to be included upon acceptance of participation. Homogenous sampling was employed by using a small pool of study participants and this meant that statistical interferences were not used to answer the research objectives and while generalizations could still be made from non-probability sampling regarding the research questions, they were not based on statistical grounds (Saunders, Lewis and Thornhill, 2015). Although there were elements of quantitative research incorporated as part of the survey processes, they functioned to support the qualitative research and do not statistically represent the general population. Participants were geographically dispersed, with most participants being based in Ireland and one participant being based in the United Kingdom.

BCT was predicted to be generally unheard of in the clinical research industry, and so the probability of most clinical research staff being aware of the technology was to be expectantly low. Those involved directly with clinical data management were the most likely candidates to have a pre-existing knowledge of BCT due to the nature of their profession and therefore represent the best-fit participants to meet the study objectives. The researcher used personal judgment based upon findings in the literature review and from real-world knowledge to come to the conclusion that those involved with clinical data management would be the most suitable candidates of having some prior knowledge of BC as BC would have a somewhat strong reputation in technological data-based professional backgrounds. Therefore, homogenous sampling was selected based upon these judgments. Selecting participants such as clinical nurses for example, would not represent a suitable sample selection as they would most likely have less in-depth knowledge regarding clinical data integrity and BC as those directly involved with clinical data management. Incorporating homogenous sampling does mean that the likelihood of the sample being representative is low and this represents a limitation of the study. Saunders, Lewis and Thornhill (2015) suggest a sampling size of between 4-12 when the nature of the study consists of a homogenous population. Given the time constraints of the study the sampling size was kept to a minimum of 9.

3.5 Reliability and Validity

Reliability incurs that a study is consistently reproducible (Saunders, Lewis and Thornhill, 2015). There are many possible issues concerning reliability when carrying out a mixed survey approach methodology. Participants can be affected by a range of factors which may interfere with their willingness or their assiduity when engaging with research studies such as surveys. Participants may rush through the questions as they may have undertaken the survey during a time-sensitive period of their day, and this may reflect a disparity in answers then if the participant undertook the survey when they had more time to complete it (Saunders, Lewis and Thornhill, 2015). While this may not necessarily be the case for quantitative aspects of the survey, it may be the case for qualitative aspects. To enhance the reliability the sampling size

could be increased, however, due to time limitations, the sampling size was kept low. Despite extenuating factors that may affect participants, survey questions were designed to produce reproducible results thereby incurring reliability with the intended study population. Reliability issues arise when using self-completed surveys as data collection methods as they can sometimes result in uninformed responses. This may occur due to the participant possessing a lack of knowledge or experience to effectively answer the questions, which can result in guesswork. However, participants are more unlikely in a self-completed survey to answer questions to please the researcher or as a result of social pressures as can happen with interview processes which can lead to increased reliability (Saunders, Lewis and Thornhill, 2015). Questions were designed to be as unambiguous as possible such that the respondent dissected and answered the question as was intended meaning the answers could be understood and interpreted effectively.

Validity which refers to the way in which the measures used as part of the research actually work to answer the research questions (Saunders, Lewis and Thornhill, 2015). Questions utilized as part of this research were designed to in the best way possible to ensure that their importance measured up to the research objectives. Both quantitative and qualitative questions within the survey have either a direct, or leading link towards answering the research questions overall.

3.5.1 Pilot Testing Survey and Alterations

While originally intended as a purely qualitative survey designed to be carried out in an interview setting, extensive alterations were made to various questions while converting some questions to quantitative multiple-choice questions. This change in data collection strategy was deemed necessary following the lack of respondents willing to participate in a purely qualitative interview-based survey. Upon altering the data collection strategy to include both qualitative and quantitative methods, participants were much more willing to complete the survey. Prior to sending the online survey for data collection, it was pilot tested with a past MSc student. Following this, minor alterations were made to some questions to improve clarity and the sequences of questions was altered to improve the logical flow of the questioning.

3.6 Analysis of Data

As the online survey consisted of a mixed method approach, data produced was both quantitative and qualitative which was designed to be complementary. These two methods were designed to support each and the two sets of data were analyzed in conjunction with each other. Thematic analysis was utilized to analyze the qualitative responses from participants. This approach is useful for identifying and analyzing themes that arise within data (Saunders, Lewis and Thornhill, 2015). In this case a theme represents an important element within the data that

relates to the overall research objectives. A theme is subject to the researcher's perception and does not necessarily depend upon prevalence of occurrence. Additionally, a theme is not necessarily required to be quantifiable, but can depend upon the researcher's deemed importance in relation to the research objectives. As themes emerge, they are coded. In the case of this research inductive coding was used which coincides with exploratory nature of the research. Inductive coding is useful when there conducting research when there is limited knowledge of the topic (Braun and Clarke, 2006). Thematic analysis was applied to the qualitative data and analyzed together with the quantitative data in order to support one another.

3.7 Access and Ethical Issues

As this research includes a literature review of data fraud and misconduct, ethical considerations are paramount to deliberate for the research process. Research questions were designed to infringe as little as possible into morally ambiguous lines of enquiries. Study participants were not questioned in such a way that would implicate themselves, their organization or any other person/organization in an unfavorable or questionable manner. Furthermore, there will be no reference made to any persons/organizations/companies. The rights of all participants should be recognized, and their dignity respected while the researcher must practice integrity and avoid misrepresenting the data and findings and avoid deception and dishonesty. Furthermore, the privacy of the participants was guaranteed including their anonymity, while their right to withdraw for any reason was respected and not challenged in any way (Saunders, Lewis and Thornhill, 2015). Each participant was briefed on the overall objective of the research and the questions that would be included and was informed that all data collected would remain anonymous and no personal information would be included or shared under any circumstances. The invitation text sent to participants for the online survey can be found under appendix 1. As this was an online survey, informed consent was built into the survey. An outline of the informed consent can be found under appendix 2. As the methodology was originally designed to consist of qualitative interviews the text invite for this is also included under appendix 3.

Chapter 4

Findings and Discussion

4.1 Topic Outline

This primary objectives of this research were to highlight and explore the key areas that affected data integrity in clinical trial research. Following this, the technology Blockchain was explored through secondary research to examine its potential appropriateness and suitability for use within clinical trial research, particularly in addressing some of the key concerns of clinical data integrity. Through secondary research it was deemed that BC such was a relatively nascent and emerging technology within clinical research that it was necessary to evaluate the awareness and perceptions of this "trust-less" technology within clinical trial data management and how it could potentially impact upon the integrity of clinical data. An online survey method approach was used to collect primary data and incorporated both quantitative and qualitative questioning styles. The answers received through the survey provided an insight into clinical data issues that participants felt required improvements and their shared perception and understanding of BCT was discovered. The presented data in this section follows the flow and sequence of the questions as they were presented in the survey to participants.

Participants were first asked a descriptive question about their current/past roles within clinical data management to assess the relevancy and validity of their responses. The question put to participants was "Are you currently involved with clinical trials and could you briefly explain your role within clinical trial data management"? Of the responses, one was not involved in any way with clinical trials but had some knowledge of clinical data management systems but worked in an IT background with a strong knowledge of BC and so was included in the results. Other roles included: Clinical Data Analysts, Clinical Data Managers, Clinical Data Coordinators, Clinical Data Leads and a Clinical Trial Manager of global studies who had strong interests in BC and Artificial Intelligence technologies. One further participant was involved in systems validations projects for clinical trial management computer systems. While these roles vary slightly, they were considered to be non-probability and homogenous.

4.2 Clinical Data Integrity Issues

The first half of the survey aimed to establish the awareness among participants in relation to clinical data integrity concerns, and to gain a deeper understanding to some of the primary issues that contribute to the integrity of the data directly from participants.

4.2.1 Clinical Data Awareness

From secondary research it was suggested that issues surrounding clinical data integrity were potentially under-recorded and under-reported. It was therefore important to first gain an insight from participants into whether they were aware that there have been ongoing issues concerning clinical data integrity before asking participants to expand deeper into the issues. This was

presented to participants quantitatively as “Are you aware that there have been ongoing concerns relating to the integrity of clinical research data?”, the results of which are outlined in figure 5.

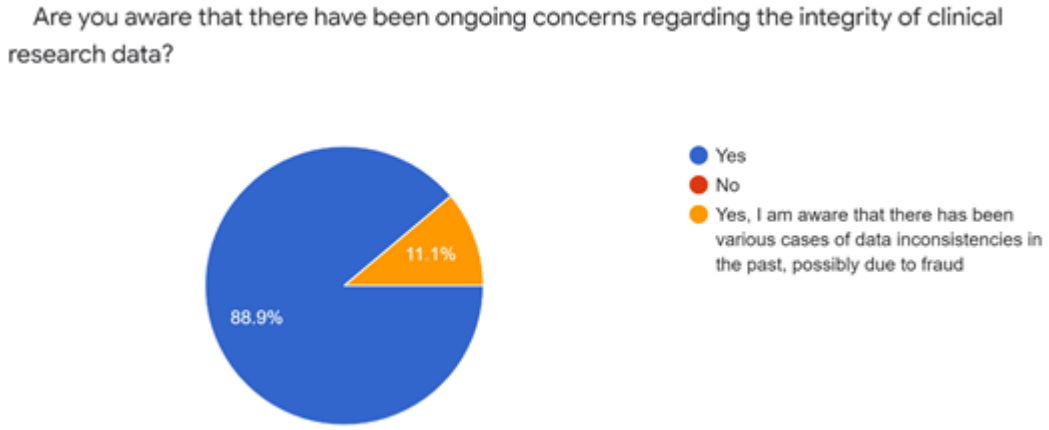

Figure 5: Pie chart representing the responses from participants when asked if they were aware of ongoing concerns with clinical data integrity

All respondents claimed that they were aware of ongoing concerns. One participant added that they were aware of data inconsistencies possibly due to fraud. This is particularly interesting as data misconduct and data fraud were not presented within the questioning in the survey, but as found in the literature, fraud and misconduct are cited to be an occurring phenomenon. The results from this section were important in order to establish the validity and relevance of the research. The responses indicate the broad awareness from the participants involved with clinical data management that there are in fact clinical data issues that are ongoing that may need addressing and highlight the need for possible data management solutions.

To expand upon this, participants were asked the qualitative question “In your opinion, is clinical data integrity an important aspect of clinical research? Please provide reasoning for your answer”. Participants elaborated on their understanding of clinical data integrity by pointing out that results gained from trials should be accurate, trusted and non-biased. Trust in data was the over-arching theme that surfaced among participants with respondents elaborating that data should be trusted to be consistent ensuring the validity of the trial’s outcome and its reproducibility. Trial reproducibility was a factor that surfaced within the literature and depends on the data being traceable, consistent and reliable. This question highlighted the understanding participants had surrounding clinical data integrity based upon their professional experience and knowledge.

4.2.2 Prominent Clinical Data Concerns

Participants were asked the quantitative question "What would you consider to be the most prominent clinical data concerns from the following?", the results of which are outlined in figure 6.

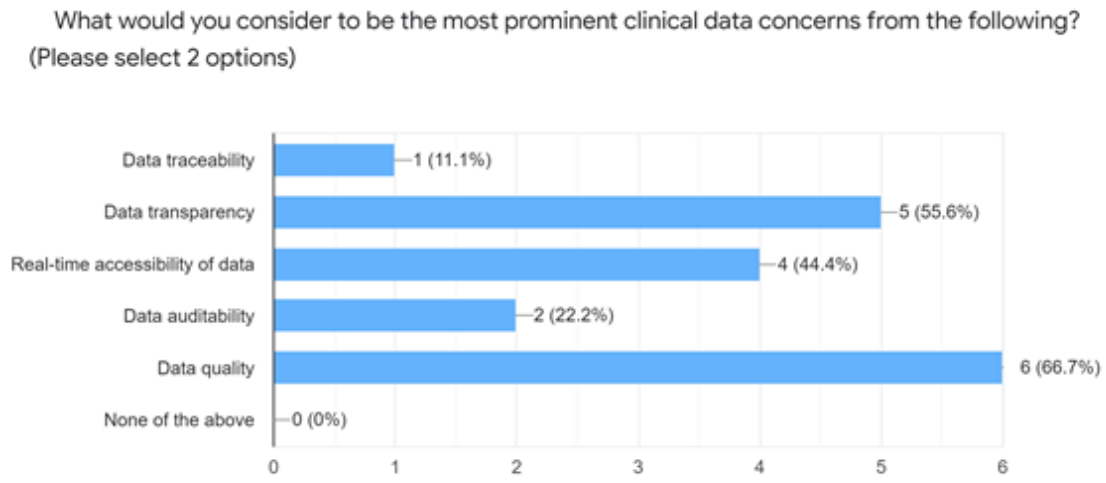

Figure 6: Bar chart representing responses participants made when asked to choose two predominant areas of data concerns

The optional choices for this question were gathered from discoveries within the literature that made up the primary general areas in which clinical data was cited to be an issue. Data transparency and data quality represented the majority of votes, followed by real-time accessibility, traceability and auditability. Data quality surfaced as a recurring theme throughout the qualitative thematic analysis as being a fundamental factor of consideration for improvement. One of the noted limitations of figure 6 was the inclusion of data quality as this could arguably be an umbrella term to include areas such as transparency and traceability. Also, to gain a more meaningful data set it may have been necessary to widen the number of participants. However, the intent of this question was to complement the qualitative responses gained from participants. It may have been more effective to exclude data quality and to ask participants to choose only one option. However, the disparate choices made by participants correlate with the findings from the secondary research as all options were chosen at least once by participants. This indicates there are multiple perceived issues that contribute to the overall integrity of clinical data.

Following this a qualitative question was asked which was "If you are aware of any clinical data areas that could be improved upon, could you outline them below and give a brief description as to why you chose them?". Responses were quite varied but supported the quantitative data gained outlined in figure 6. One participant pointed out that there are frequent changes in clinical teams

which can lead to increased loss of trial knowledge, however, did not elaborate further leading to a degree of ambiguity in the interpretation of the answer. However, another participant also contributed, that due to a lack of clinical personnel understanding, maintaining different sources of data could prove quite challenging. They pointed out how maintaining various sources and matching within clinical databases could prove quite difficult when personnel do not understand the requirements. A further participant also mentioned how securing sources of data can be problematic and that there can be incompatibilities between paper and digital records. Two separate participants mentioned the need for trust in clinical data within vaccination data management with one participant mentioning the possibility of a central repository for which all vaccination clinical data originated from for which the data the integrity of the data could be guaranteed. This presents a very applicable use for BC as BC is designed to be a single source of all data that by design ensures both privacy of data and integrity through its immutability design functionalities.

4.2.3 Real-Time Accessibility to Clinical Data

It was difficult to ascertain the level in which clinical data staff had access to real-time clinical data from secondary research. To gain an insight into this, participants were asked "In your place of work is there real-time access to all clinical data as it becomes available?", the results of which are outlined in figure 7.

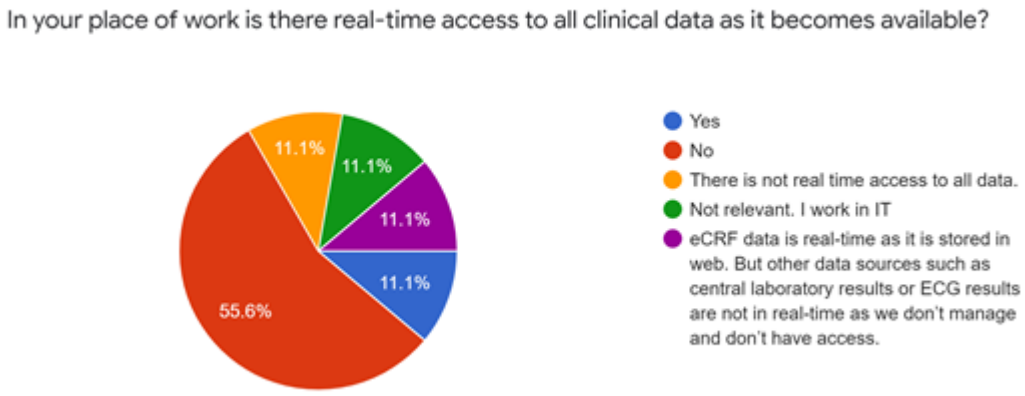

Figure 7: Pie chart representing responses when participants were asked if they had real-time accessibility to all clinical data

Six of the respondents claimed they did not have real-time access. An "other" option was included in which one respondent stated that they had partial real-time access to certain data such as data from electronic Case Report Forms (eCRF) but did not have real-time access to central laboratory or electrocardiograph results. One respondent stated they did have real-time access but did not elaborate further and one respondent chose irrelevant as they work in IT. One of the key

characteristics of BCT is its architectural ability to showcase data in full-time allowing all relevant parties access to data as it appears and so it was interesting to find that even in today's data management systems, real-time accessibility to all clinical data is often not available. This creates a strong case for the implementation of BC with clinical data management.

Following this a qualitative question was asked which was "In your opinion, would clinical trials benefit from real-time access to data as it becomes available? Please provide an explanation for your answer". All respondents emphasised the benefit from having access to clinical data in real-time. Two participants pointed out how sponsors would benefit from this as they often require data in time manners but as one participant pointed out that the collation of the data takes before it can be sent. Participants also pointed out how following up issues would be made substantially easier with real-time access with two participants pointing out how better decisions regarding data could be made more efficiently resulting in the easing of the trials progression while enhancing the quality of the data. Two participants also mentioned the potential of real-time access for vaccination trials with one participant mentioning the increased trust a patient would be provided with if they knew the integrity of the data was assured. While another participant mentioned the possibility of improved delivery times of vaccinations if real-time access was implemented. While real-time accessibility of data was not stated to be among the top two primary concerns as seen in figure 6, it can still be considered an important aspect that could improve the integrity of clinical data. Having real-time access to data allows erroneous and inconsistent data to be identified at the time of origin rather than down the line.

4.2.4 Legacy Clinical Data Management Robustness

Participants were "How robust, in your opinion are current Clinical Data Management systems in ensuring clinical data is accurate, traceable and intact?", the results of which are outlined in figure 8.

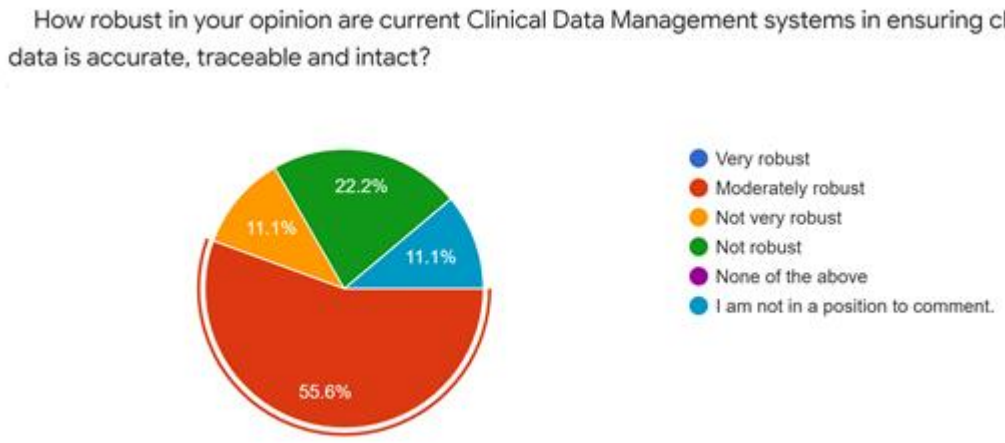

Figure 8: Pie chart representing the responses participants made when asked how robust Clinical Data Management systems were in ensuring clinical data is accurate, traceable and intact

Participants were asked this question to gain an idea of how effective they perceived their current data management systems to be in contributing towards clinical data integrity. Responses greatly varied with 5 of the respondents stating that current clinical data management systems were moderately robust while 2 of the respondents claimed they were not robust. This question was critically lacking in relevance and failed to allow for a deeper analysis into current clinical data management systems. Although this question lacked relevancy and depth it did however show the rather disparate opinions participants shared in relation to their current data management systems.

4.3 Blockchain and Clinical Data Management

Following establishment of the primary clinical data concerns from respondents, they were then asked questions pertaining to BC. These questions established the awareness, understanding and perceptions of the technology as well as highlight the obstacles and challenges facing its implementation.

4.3.1 Blockchain Awareness

Part of this research was to determine the level of awareness among participants and so participants were asked the question "Prior to this study, were you aware of the technology called Blockchain?", the results for which are outlined in figure 9.

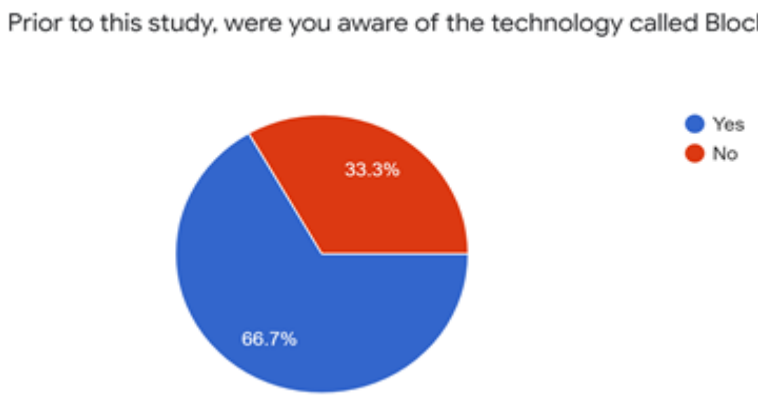

Figure 9: Pie chart representing the responses from participants when asked about prior knowledge of BC

Participants answers relating to subsequent BCT questions were dependent upon their prior knowledge and awareness of the technology, so it was positive to find that more than half were aware of BC prior to starting the survey. It was found that 6 participants possessed prior awareness while 3 stated they were not aware prior to the commencement of the survey. Participants were purposively chosen who best represented the professions in which they would most likely have heard of BCT. As BC is currently being implemented across a range of industries for purposes such as data management and traceability, it was assumed that those working in clinical data

management would have some prior knowledge of the technology based upon their professional backgrounds. However, the value was expected to be lower and it was a concern that if the overall level of awareness was lower that it would impact or possibly invalidate the remaining responses regarding BC and its role within clinical research. Of those who stated they were not previously aware of BCT involved those involved with clinical data coordination, clinical data management and data governance and a participant whose profession was clinical data lead. It is interesting to note that in the introduction of chapter 1, it was found that in 2018, out of 64 PubMed publications with the word 'Blockchain' in the title, only 4% were represented by the area of clinical trials. While it was found that interest with BC within clinical research was accelerating, awareness within the industry was still relatively low, and while the results of this study are not intended to be empirically representative of a population, it is interesting however to note that participants such as those involved with clinical data coordination and clinical data management and governance, had not previously heard of this technology. This does possibly highlight the level in which BC is still in its incipient stages of application within the industry.

4.3.2 Blockchain Understanding

Participants were also asked "How would you describe your understanding of Blockchain?", the results for which are outlined in figure 10.

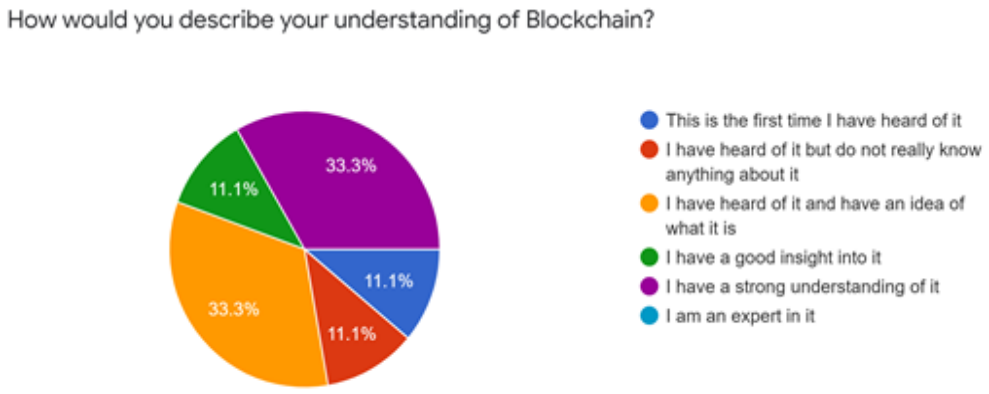

Figure 10: Pie-chart representing the responses participants made when asked to rate their understanding of BC

Responses greatly varied, however it was interesting to find that 3 respondents claimed to have a strong understanding of BCT, while 3 claimed to have an idea of what the technology was. This quantitative response does not correspond with when participants were asked if they had prior knowledge of BC before commencing the survey. Previously, 3 participants had mentioned that

they had no prior knowledge of the technology which does not coincide with the previous results where only 1 of the respondents stated it was the first time they had heard of BC. This does suggest a degree of unreliability with the survey and it is possible questions were not presented as clear and concise as they could have been. It is also possible there was not enough differentiation created between these two questions regarding awareness and understanding of BCT. Gaining an insight into participants awareness and understanding of BCT as well as their awareness of clinical data integrity concerns was important to identify themes that surfaced when qualitatively asking participants further questions. It was to be expected that there would be a degree of a lack of prior knowledge of BC but those participants who had not heard of BC were still able to add meaningful data relating to clinical data integrity issues. As participants were chosen who best represent the population that would most likely have pre-existing knowledge of the technology based upon their professional backgrounds, it can be surmised that the awareness and knowledge extending beyond clinical data management within the clinical research industry is most likely to be lower.

4.3.3 Impact of Blockchain on Clinical Data Integrity

Participants were then asked the qualitative question "Given your prior knowledge of (or from the introduction provided) of Blockchain, how do you feel clinical data integrity could be impacted by this technology?". This provided an opportunity for participants give their insight and opinions into how BC could benefit clinical data management. Responses were quite mixed with one respondent simply stating that it could improve traceability, transparency and integrity while a second respondent claimed it could improve the quality, integrity and reliability of clinical data. One participant pointed out that there exists a degree of traceability by audit trails with datapoints integrated in current management systems but that BC they believed would enhance that that traceability. Another participant mentioned how BC would improve the audit trail for any data entered into the study. Another respondent pointed out the benefit BC provides regulatory authorities and future researchers as having all data on a BC would aid in the reproducibility of the trial.

4.3.4 Challenges Facing Blockchain

Participants were asked the question "In your opinion what do you think could be the primary challenges in adopting this technology within clinical trial research?". Participants were asked to choose 2 options and the results are outlined in figure 11.

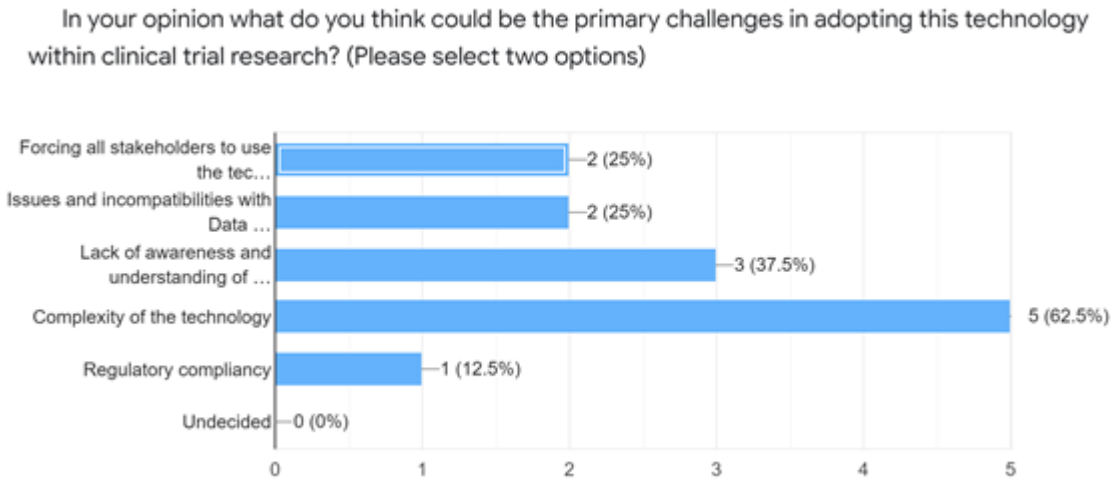

Figure 11: Bar chart representing choices participants made when asked to choose two options that would likely present as challenges when considering BC adoption within clinical research

The responses were as follows: Forcing all stakeholders to use the technology - 2. Issues and incompatibilities with data protection laws such as GDPR – 2. Lack of awareness and understanding of the technology - 3. Complexity of the technology - 5. Regulatory compliancy - 1. While it was not part of the research objectives to examine the challenges associated with BC and clinical research, it was appropriate in determining the suitability of its implementation. Complexity of BC scored was rated the highest and considering the responses from participants, this was not particularly surprising. Following this was a lack of awareness and understanding of the technology. It is interesting to note that complexity of the technology was the highest rated challenge associated with BC adoption. Despite this being rated as the highest challenge, the majority of respondents claimed they would be inclined to use BC over current legacy data management systems. However, even so participants have recognised that BC is still a complex technology. One participant pointed out that there needs to be BC experts who are proficient in areas of BC and clinical trial data management. As it stands, this is an area that is lacking as BC is an inherently complex piece of software architecture. An interesting avenue to explore would be the user-friendliness of its user interface in managing clinical data. This would however be a test limited to actual testing of the technology within a clinical trial management environment. It was not surprising that complexity and also lack of awareness and understanding of BC rated the highest obstacle. Participants within the survey most likely would not have prior knowledge or experience when concerned with for example the incompatibilities of BC with data protection laws or issues with regulatory compliancy. It seems likely that the lack of understanding and

complexity need to be addressed before participants can assert valuable responses in relation to other areas of challenges.

Following this, participants were asked to the qualitative question "In your opinion do you think stakeholders such as regulatory authorities and clinical researchers would benefit from sharing and storing clinical data on a Blockchain? Provide reasoning for your answer". All responses were favourable towards the benefit stakeholders would gain from using BC. One participant pointed out humanity have produced more data in the last 10 years than the entire history of our species previous to that. They further mentioned how the production of data is increasing exponentially every day. It can only be assumed that as we progress into the future, we will continue to increase the production of our clinical data continuously. It is therefore reasonable to suggest that there should be suitable way to manage that data to ensure its complete traceability and integrity. For clinical research, at this point in time BC appears in theory to be able to provide that. Participants mentioned how clinical researchers would have access to data contained in a BC to further their studies while resting assured that all data is immutable and can be proven to have not been tampered with in any way.

4.3.5 Reputation and Perceptions of Blockchain

Participants were asked a qualitative question which was "Blockchain is typically associated with Cryptocurrencies such as Bitcoin and negatively perceived online activities. How do you feel this would impact upon the perceptions and adoption of this technology within clinical research"? BC is a platform that was built and originally intended for Bitcoin. As was seen in the literature review the perceptions of Bitcoin are rather ambivalent and therefore it was prudent to gain an insight into participants perceptions and attitudes to BC. The central idea of this research revolves around data integrity and trust. BC by design is a "trust-less" technology meaning the need for trust is designed to be omitted from its functionality. However, the pharmaceutical industry is particularly conservative as mentioned by one participant and when personal data such as clinical data is considered there may be a degree of distrust placed in a new data management platform. The same participant pointed out that the conservativeness of the clinical research industry can be based upon the uncertainties of whether a drug is safe and effective yet and therefore adopting BC will require extensive understanding from different stakeholders. Two participants mentioned how they think of cryptocurrencies when hearing BC which could weaken its brand and present uncertainties in that people may be suspicious of sharing their personal details with BC. Another participant added that cryptocurrency transactions within BC are trusted even though they take place in an un-trusting environment. They mentioned that this professes proof of its applicatory purposes. It is ironic that that there are trust issues associated with using the technology, despite the technology being designed to be trust-free. Responses to the perceptions were mixed, while some mentioned that people will be suspicious of sharing personal data which is out of their

control which was consolidated by a further response mentioning security issues with protecting patient data while a further added that dissociating BC from Bitcoin may prove difficult. From the responses gathered it is evident there is a degree of mistrust and uncertainty of the technology, however this may come down to a lack of understanding as 3 participants claimed to have little to no understanding of the technology. Adoption of new technologies can only really be considered if they are accepted and understood by those who will be using them.

4.3.6 Is BC 'Over-Hyped'?

Participants were asked the question "It has been suggested that Blockchain is an 'over-hyped technology. Given your knowledge, how would you respond to this statement?", the results of which are outlined in figure 12.

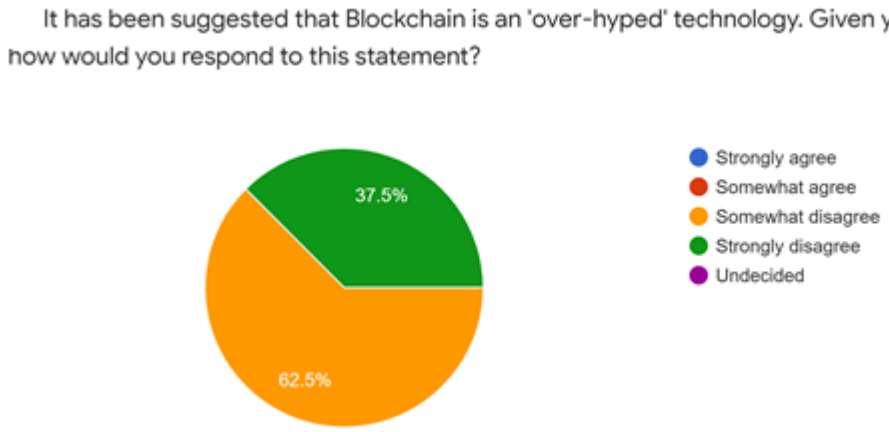

Figure 12: Pie-chart representing the choices participants made when asked to rate their response to the statement of BC being an 'over-hyped' technology

It had been suggested from secondary research that BC was facing a stagnation of innovation plateau, in that it was facing the issue in which many new technologies face – being 'over-hyped'. Participants were asked this as it ties in with the research objective of examining the perceptions of the technology. Surprisingly responses fell between 5 respondents choosing they somewhat disagree and 3 choosing they strongly disagree. Again, there are disproportionate inconsistencies in the data here as participants who stated previously that they had no prior knowledge of BCT should have been more inclined to choose the undecided option. This can incur several things; some participants made an uninformed choice, or the survey was biased towards the positive characteristics of BC introduced by the style of questioning. This questions the validity of the questions and again questions the reliability of the survey. Although certain elements of data may be skewed, 3 of the participants chose to strongly disagree with BC being an 'over-hyped' technology. This indicates that perceptions among participants is in favour of the technology, but it is difficult to ascertain a true value due to the inconsistencies of data from certain participants.

It should also be noted that one participant possibly chose not to answer this question as they previously stated they had no prior knowledge of BC.

Finally, participants were asked "Would you personally be inclined to use this technology over current legacy data management systems?", the results of which can be seen in figure 13.

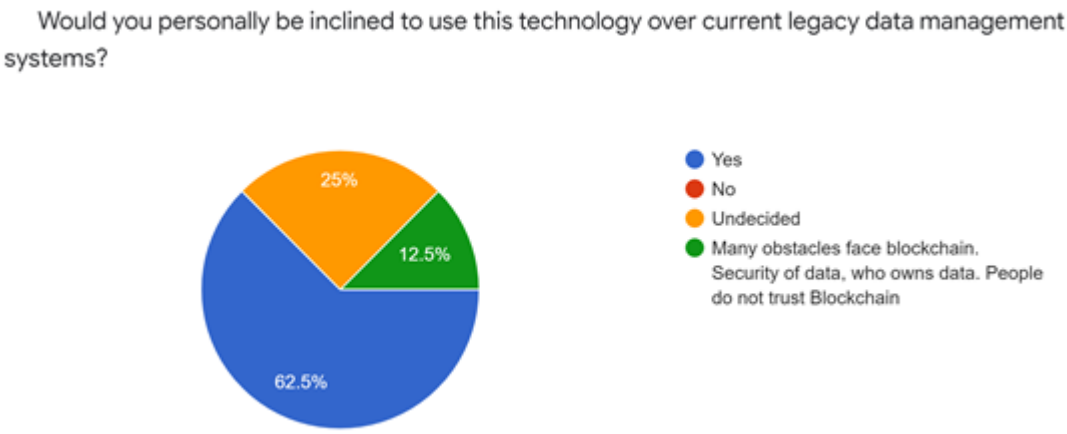

Figure 13: Pie chart representing responses from participants when asked if they would be inclined to use BC over current data management systems

The majority of respondents claimed they would be inclined to use BC over current legacy data management systems which is particularly interesting given the varied responses to previous questions such how well respondents understood the technology. In figure 10, 3 respondents claimed to have a strong understanding of the technology while 3 claimed to have an idea of what it was while the remaining stated, they did not know anything about BC. It should be noted that one respondent chose not to answer this question. One respondent chose to mention the obstacles facing BC but did not indicate if their inclination towards adopting the technology. It is interesting to note that while a degree of uncertainty persisted in participant responses, most stated they would be inclined to use BC which is promising.

4.4 Themes Identified

Thematic analysis was applied to identify the common recurring themes that surfaced from qualitative responses. Some of the primary themes are outlined below.

4.4.1 The Patients Role

A theme that surfaced within the research was the role of the patient and how they could benefit from a BC-enabled clinical trial. The patient's role was something that was gathered from carrying out secondary research and although patients were not included in the survey, responses indicated that patients are more involved with their personal health and health-related data. One respondent

suggested that data contained on a BC that is immutable and encrypted would enhance the level of trust patients have in sharing and entrusting their data with clinical researchers. The patient is particularly important to consider as they are arguably the most important stakeholders within clinical trials. Patients who participate in clinical trials are often relying on the success of the trial to treat their specific medical needs and it is necessary that they trust that their data is secure, safe and traceable.

4.4.2 Trust in Clinical Data

Trust in data is a theme that popped up quite frequently throughout the survey responses. Almost all respondents in some form mentioned trust in data. One participant stated that the data collected in a clinical trial should be exactly as it was, that it has not been tampered with and there should be a degree of confidence that it has not been altered in any way especially when it forms the basis for clinical research. Another respondent also added that clinical trials are responsible for producing high quality data and it is critically important that the data remains unaltered and stored in a secure environment. This was a fundamental theme that also surfaced when carrying out secondary research as it was found that data fraud and misconduct had occurred throughout trials in the past affecting the trust placed in the data. Trust is an overarching factor when considering any kind of data, especially clinical data. Fundamentally there must be trust that the data is accurate and truly representative of the trials results, without trust, reliability and reproducibility of a trials methodologies and results are called into question. The patient must be able to trust their data is in safe hands, the regulators must be able to trust that the data is accurate, transparent and auditable. These are all elements built into BC technology. Although BC cannot prevent erroneous data at the point of origin, in other words it cannot prevent someone inputting wrong data, it can however provide a trail of time-stamped connected data from the point of origin right throughout the entire data production process. Data cannot be altered or changed in the append-only ledger, thereby incurring trust in that the data has not been modified or altered at any point.

4.4.3 Trial Sharing of Data

A theme that cropped up was the sharing of clinical data between ongoing trials. This particular area was not researched as part of the study. However, as one respondent mentioned having a central repository of depersonalised clinical data would be shared within the clinical research industry would be particularly beneficial. While the idea of this is attractive, there may be numerous complications of achieving a central repository of clinical data such as the ownership of data. Another respondent stressed the need to keep all personal clinical data from appearing within data management systems. The respondent mentioned this as an area that required improvement and that personal data should never appear in data systems. This ties in with the need to create trust in the storage and sharing of personal clinical data as mentioned previously. The respondent went on to mention how there is a huge amount of data circulating the world

today and that more data has been created in the last 10 years than the entire history of the human species. This is a particularly interesting point to take into consideration to take into account when considering the implementation of BCT. BCT, according to the research has the architectural capabilities to store exponential amounts of data in a distributed shared ledger, while maintaining the data in an encrypted and immutable format.

4.4.4 Vaccination Trials

The theme of vaccinations surfaced from responses. Some participants pointed out the need for integrity in clinical data, especially in relation to vaccination data. The theme of vaccinations is a particularly interesting area to examine in relation to BC. Participants noted that how the compilation of vaccination data across various trials could be stored in a central repository such as BC and that real value in data could be harnessed by combining artificial intelligence pattern recognition software with the data produced from numerous trials. This is an attractive concept, however, the practicalities of this would face enormous obstacles. The question of ownership of and accessibility of the data would be called into question. In theory, having all clinical data from all clinical trials across multiple pharmaceutical companies stored and shared in a single master BC would be enormously beneficial to the advancement of medicine, however, is an unrealistic expectation. From the responses it can be indicated that there is a degree of mistrust in vaccination clinical data and this is potentially consolidated due to the Covid-19 pandemic that is currently ongoing at the time of this study. One participant mentioned how BC could be used to improve the trust from a patient perspective for vaccination trials while another participant mentioned how delivery times would be increased.

4.5 Discussion

When considering clinical data integrity, it necessary to consider how data transparency, traceability, real-time accessibility, auditability and data quality all contribute to the overall integrity of clinical data. These areas were all identified as concerns among participants through both quantitative and qualitative questioning and these correlated with the findings from the secondary research. Themes that were discovered included how trust could be enhanced for patients, the role of real-time data accessibility within vaccine trials, how BC could improve the integrity of clinical data by preventing tempering and alteration of data once uploaded. The general consensus gained from participants indicated a favourable perception towards BC however, it was clear from participants that there exist various of challenges to overcome. Complexity of the technology and a lack of awareness of the technology represented the highest quantitative responses from participants as challenges. As Bitcoin and consequently Blockchain may have unfavorable public reputations, this may act as a challenge for the uptake of the technology in traditionally conservative industries such as the pharmaceutical industry. Lack of public trust in a designed to be “trustless” technology may ironically act as a prohibitory factor in

its widespread adoption within clinical research, however this remains to be seen. Responses to questions were quite varied and a degree of inconsistency was noted throughout the survey responses. There are clearly many potential uses for a technology such as BC. Its application for real-time sharing of vaccination trial data is a particularly interesting area of potential future research as two participants seemed confident in the need to improve vaccine data trust and delivery times of vaccines.

The first research objective of this research was:

- To gain a deeper understanding of the some of the key issues surrounding clinical research data integrity

Primary data gained from this study presented the areas of concern that were to be expected as proposed by the secondary research. However, limitations of the scope of questions and the lack of follow up investigation prevented a thorough in-depth analysis into these specific areas. While the key issues were identified, their root causes why they occur were not identified. Participants did not elaborate much beyond the specific issues that had been identified. However, some participants did point out incompatibilities and issues between digital and paper records which is an area that could be further explored in future research.

The second research objective was:

- To evaluate the current knowledge and awareness of Blockchain within the clinical research industry

The primary data gained from this research pertaining to this objective was mixed. There were inconsistencies in responses throughout the survey, so it was difficult to ascertain a true representative idea of the awareness and knowledge of BC among participants. It was however, predicted that the majority of participants would have no prior knowledge, but it was found that more than half had at least heard of BC prior to the study. It was also found that 3 participants had a strong understanding of the technology, 1 had a good insight into it while 3 had an idea of the technology. It is clear from this research that BC has a long way to before it recognized as a potential data management solution in clinical research. The limited survey approach and limited number of participants reduced the capacity to enable a conclusive answer to meet this research objective.

The final research objective was:

- To determine the suitability of Blockchain in addressing the key issues surrounding clinical data integrity from a clinical data management perspective through quantitative and qualitative research

This objective represented the primary research outcome. The general consensus gleaned from responses were generally favorable towards the application of BC. Respondents stated how real-time accessibility that BC offers could significantly improve the integrity of data due to the immutability and traceability that BC offers. Participants generally stated how stakeholders, including the patients would benefit from having BC-enabled clinical trials. BC posses the architectural structure to meet the numerous concerns that were found from both primary and secondary research. However, there is much research and studies to be conducted in the future to determine how BC can overcome the numerous challenges that stand in the way of its implementation.

Chapter 5
Conclusions

5.1 Conclusion

A mixed methods approach was conducted as part of this research study to examine the impact of BC on clinical trial data integrity. The first objective of this research was to delve deeper into the predominant clinical data issues that are cited to be ubiquitous throughout the clinical research industry. Clinical data issues such as quality, transparency, traceability, auditability and real-time accessibility to data were all issues that were regularly encountered throughout the secondary research. Through thematic analysis from qualitatively gained data, it was found that most of these issues were mentioned by participants as being areas that could be improved upon. Correlation with participant responses and what was found in the literature relating to clinical data integrity concerns shared similarities. It can be stated that all of these particular areas within data management all collectively contribute to the level of integrity and trust that can be harnessed from clinical data. Data is such an important aspect to modern day society that there must be a foundation of trust embedded within data and this is especially relevant to personal health data. Clinical trials embody a fundamental aspect of modern-day medicine and it is fundamental that the data produced from them is completely representative of the trial so that patients receive the necessary care and treatment they require. While primary research confirmed the existence of these issues, it did not highlight why or how these issues arise, however despite this, primary data gave an insight into some of the issues that occur and highlighted areas of primary concern such as data quality and transparency.

It is evident from primary research that BC has a long way to go in terms of awareness and understanding of the technology as was expected. There still remains a great deal of ambiguity surrounding the technology such as its association with cryptocurrencies while there are numerous challenges to overcome such as increasing the perceptiveness of clinical researchers and patients to the technology's potential and this represents one of the conclusions from the research. Without actually testing BCT within clinical trials it is not possible to say from this research that it will solve the extensive clinical data integrity-related issues encountered in the clinical industry. However, it can be concluded from the research that there is general favour for its potential applicatory benefits but there are many challenges and obstacles the technology must first overcome. The final research objective was to determine the suitability of Blockchain in addressing the key issues surrounding clinical data integrity from a clinical data management perspective Until these challenges are addressed it not possible to conclude the final research objective. However, until these Blockchain-related challenges are recognised, addressed and solved it is not possible to conclusively state that BC is suitable for addressing the clinical data integrity concerns.

In general, the purpose of the research was met. Clinical data issues were identified through secondary research and correlated with the primary research of this study, awareness and

understanding of BC by clinical research data management professionals was assessed and the suitability of BC was recognized from both primary and secondary research. The purpose of this study can be considered justified as BC has all the fundamental characteristics built into its design to meet the demands required to improve the integrity of data produced throughout clinical trials. The trust placed in clinical data is fundamental to the advancement of safe medicine for patients and this is true now more than ever as clinical data is being produced in vast quantities spread across multiple complex data streams that will only increase over time. BC not only has the potential to improve the integrity of the data but can also build trust back into clinical research. There is significant further test studies and research to be carried out opening up vast areas of potential future research for its applicability within clinical research.

These are the final conclusions of the study:

- There are extensive data-related issues within clinical research that contribute to the lack of integrity in the resulting data
- There are many potential uses for Blockchain within clinical trials beyond addressing integrity
- Blockchain could be used to improve the trust patients have in enrolling within clinical trials and sharing their data with clinical researchers
- There is a slight lack of awareness and understanding of Blockchain which could inhibit its implementation
- There are numerous challenges currently inhibiting its application within clinical research and these need to be addressed
- If BC overcome the challenges facing it, it could overhaul how clinical data is shared, managed and stored and could usher in an era where clinical data is "trustless"

5.2 Limitations of the Study

Following completion of this study there were various alternative directions the researcher could have taken to answer the research objectives. As mentioned in chapter 3 of this research, it was the researcher's intent to carry out qualitative interviews to discuss first-hand the issues that impact upon clinical data integrity. For this strategy an interpretivist philosophy would have been adopted to gain qualitative data to provide deep insights into the issues at hand based upon personal experiences and industry-based knowledge in a primarily inductive nature. Given more time a purely qualitative inductive approach would be taken and following that based upon the findings a quantitative deductive survey would be dispersed resulting in a sequential exploratory

mixed methods design. This would allow emphasis to be placed upon the qualitative exploration of data integrity issues and objectives pertaining to BC and themes discovered would be used to form the quantitative research. Results from these two methods could be then be integrated and triangulated to provide further depth and understanding to the research while providing more conclusive results.

The first objective of this study was to gain a deeper understanding to some of the key issues surrounding clinical data integrity. Data gathered from secondary research suggested primary areas of concern, however implementing the survey method was limited in scope to delve deeper into these issues as it was not possible to conduct follow up questions in the self-completed survey. Conducting semi-structured interviews may have allowed for a more comprehensive analysis of the problems facing clinical data and further exploration of the implementation of the technology. Ultimately due to the methodological approach the study failed to thoroughly meet the research objectives. The sampling method was also limited as although it was not the intent to present the data representatively of the population of those involved with clinical data management, a larger more diverse sampling method would have provided more conclusive data. Participants who did not have prior knowledge of BC called into question the reliability of the research strategy and this was evident within the survey as there were inconsistencies in the data identified that were associated with BC awareness and understanding. It is therefore reasonable to presume that there were elements of uninformed decisions from certain participants and/or bias introduced in favour of BC throughout the survey that influenced participants decisions. In addition, some questions could not be considered reliable as a degree of ambiguity in responses was noted. Some questions lacked relevancy to the overall research objectives.

Upon completion of the survey process, select questions were identified that had limited scope in addressing the research objectives and provided irrelevant data. These questions could have been more integrative of BC and clinical data integrity. A further limitation was the overall scope of the research. Data integrity could be considered a broad umbrella term that encapsulates numerous data issues and it is therefore possible that the research could have been narrowed in scope to identify one particular area of interest in which to apply BCT. One further limitation was the overall scope of the research. It may have been prudent to select one of the areas of clinical data concerns such as data traceability for example and focused on how BC could be implemented to improve that specific area. Trying to address the impact on clinical data integrity overall increased the breadth of the research beyond the time allowances that were given for the research to be carried out.

5.3 Recommendations

The implementation of Blockchain technology within clinical trial research holds many potential benefits for improving the transparency, traceability and overall integrity of clinical data, however, there are numerous challenges that face the coalescence of the two. Challenges such as personal data security, compatibilities of BC with current clinical data management systems, forcing all stakeholders to use BC as well as regulatory compliancy issues all face the implantation of the technology. To expand upon this research, the challenges that face BCs application with clinical trial research should be explored in greater detail. It is evident from this research that there is an interest in the technology for meeting the issues of data integrity, however, BC also holds additional beneficial applicatory purposes for areas such as clinical trial recruitment, protocol management, real-time patient monitoring and shared clinical trial research in a central repository. Throughout the secondary research carried out as part of this research there was seldom mention of any analysis or studies carried out on the obstacles that face the application of the technology and this would be a very promising area for further research. One particular area which would be worth pursuing is how to ensure BC is compatible with GDPR while storing and sharing personal clinical data as this was a theme that arose in both primary and secondary research. Another potential area is the use of BC in sharing clinical data across multiple vaccination trials.

Bibliography

Albanese, G. *et al.* (2020) 'Dynamic consent management for clinical trials via private blockchain technology', *Journal of Ambient Intelligence and Humanized Computing*. doi: 10.1007/s12652-020-01761-1.

Alshamsi, A. and Andras, P. (2019) 'User Perception of Bitcoin Usability and Security across Novice Users'.

Bashir, I. (2017) *Mastering Blockchain: Deeper insights into decentralization, cryptography, Bitcoin, and popular Blockchain frameworks*. 2nd edn. Birmingham, UK: Packt Publishing.

Benchoufi, M., Porcher, R. and Ravaud, P. (2018) 'Blockchain protocols in clinical trials: Transparency and traceability of consent', *F1000Research*, 6. doi: 10.12688/f1000research.10531.5.

Benchoufi, M. and Ravaud, P. (2017) 'Blockchain Technology for Improving Clinical Research Quality', *Trials*, 18(1), p. 335. doi: 10.1186/s13063-017-2035-z.

Braun, V. and Clarke, V. (2006) 'Using thematic analysis in psychology', *Qualitative Research in Psychology*. Routledge, 3(2), pp. 77–101. doi: 10.1191/1478088706qp063oa.

Burrell, G. and Morgan, G. (1979) *Sociological Paradigms and Organisational Analysis: Elements of the Sociology of Corporate Life*. 1 edition. London: Pearson Education.

Carlisle, J. B. (2017) 'Data fabrication and other reasons for non-random sampling in 5087 randomised, controlled trials in anaesthetic and general medical journals', *Anaesthesia*, 72(8), pp. 944–952. doi: 10.1111/anae.13938.

ClinicalTrials.Gov (2020) *Trends, Charts, and Maps - ClinicalTrials.gov, ClinicalTrials*. Available at: https://clinicaltrials.gov/ct2/resources/trends (Accessed: 16 May 2020).

Crosby, M. (2016) 'BlockChain Technology: Beyond Bitcoin', *Applied Innovation Review*, (2), p. 16.

European Parliament (2020) *Blockchain and the General Data Protection Regulation*. Brussels: European Parliamentary Research Service, pp. 1–101. Available at: https://www.europarl.europa.eu/RegData/etudes/STUD/2019/634445/EPRS_STU%282019%29634445_EN.pdf (Accessed: 12 July 2020).

FDA (2013) 'Guidance for Industry Oversight of Clinical Investigations — A Risk-Based Approach to Monitoring'. Available at: https://www.fda.gov/media/116754/download.

FDA (2016) 'Data Integrity and Compliance With CGMP Guidance for Industry'. Food and Drug Administration. Available at: https://www.fda.gov/files/drugs/published/Data-Integrity-and-Compliance-With-Current-Good-Manufacturing-Practice-Guidance-for-Industry.pdf (Accessed: 21 June 2020).

Friar, H. S. and Kirsh, D. (2019) *Why data integrity is important for the success of clinical trials, MassDevice*. Available at: https://www.massdevice.com/why-data-integrity-is-important-for-the-success-of-clinical-trials/ (Accessed: 24 June 2020).

George, S. L. (2016) ‘Research misconduct and data fraud in clinical trials: prevalence and causal factors’, *International Journal of Clinical Oncology*, 21(1), pp. 15–21. doi: 10.1007/s10147-015-0887-3.

George, S. L. and Buyse, M. (2015) ‘Data fraud in clinical trials’, *Clinical investigation*, 5(2), pp. 161–173. doi: 10.4155/cli.14.116.

Gupta, A. (2013) ‘Fraud and misconduct in clinical research: A concern’, *Perspectives in clinical research*. (Perspectives in Clinical Research), 4(2). doi: 10.4103/2229-3485.111800.

Hirano, T. *et al.* (2020) ‘Data validation and verification using blockchain in a clinical trial for breast cancer’, p. 30.

Hölbl, M. *et al.* (2018) ‘A Systematic Review of the Use of Blockchain in Healthcare’, *Symmetry*, 10, p. 470. doi: 10.3390/sym10100470.

Holdowsky, J., Lele, N. and Lougheed, G. (2018) *Many paths lead to blockchain adoption, and no two are alike*, *Deloitte Insights*. Available at: https://www2.deloitte.com/us/en/insights/topics/understanding-blockchain-potential/global-blockchain-survey/2019-adoption-by-industry.html (Accessed: 8 July 2020).

Hume, S. *et al.* (2017) ‘Visualizing and Validating Metadata Traceability within the CDISC Standards’, *AMIA Summits on Translational Science Proceedings*, 2017, pp. 158–165.

ICH (2018) ‘E6(R2) Good Clinical Practice: Integrated Addendum to ICH E6(R1) Guidance for Industry’. International Council for Harmonization. Available at: https://www.fda.gov/media/93884/download (Accessed: 13 May 2020).

Khin, N. A. *et al.* (2020) ‘Data Integrity in Global Clinical Trials: Discussions From Joint US Food and Drug Administration and UK Medicines and Healthcare Products Regulatory Agency Good Clinical Practice Workshop’, *Clinical Pharmacology and Therapeutics*. doi: 10.1002/cpt.1794.

Krishnankutty, B. *et al.* (2012) ‘Data management in clinical research: An overview’, *Indian Journal of Pharmacology*, 44(2), pp. 168–172. doi: 10.4103/0253-7613.93842.

Kuo, T.-T., Kim, H.-E. and Ohno-Machado, L. (2017) ‘Blockchain distributed ledger technologies for biomedical and health care applications’, *Journal of the American Medical Informatics Association*. Oxford Academic, 24(6), pp. 1211–1220. doi: 10.1093/jamia/ocx068.

Lin, I.-C. and Liao, T.-C. (2017) ‘A survey of blockchain security issues and challenges’, *International Journal of Network Security*, 19(5), pp. 653–659. doi: 10.6633/IJNS.201709.19(5).01.

Mackey, T. K. *et al.* (2019) ‘“Fit-for-purpose?” – challenges and opportunities for applications of blockchain technology in the future of healthcare’, *BMC Medicine*. (Beyong Big Data to new Biomedical and Health Data Science), 17, pp. 1–17. doi: 10.1186/s12916-019-1296-7.

Mann, S. P. *et al.* (2020) ‘Blockchain, consent and prosent for medical research’, *Journal of Medical Ethics*. Institute of Medical Ethics, pp. 1–7. doi: 10.1136/medethics-2019-105963.

MarketsandMarkets (2020) *Blockchain Market Size, Growth, Trends and Forecast to 2025 | MarketsandMarkets, Blockchain Market by Component (Platform and Services), Provider (Application, Middleware, and Infrastructure), Type (Private, Public, and Hybrid), Organization Size, Application Area (BFSI, Government, IT & Telecom), and Region - Global Forecast to 2025*.

Available at: https://www.marketsandmarkets.com/Market-Reports/blockchain-technology-market-90100890.html (Accessed: 12 August 2020).

Maslove, D. M. *et al.* (2018) 'Using Blockchain Technology to Manage Clinical Trials Data: A Proof-of-Concept Study', *JMIR Medical Informatics*, 6(4). doi: 10.2196/11949.

MRC (2014) 'MRC Policy and Procedure for Investigating Allegations of Research Misconduct'. Medical Research Council. Available at: https://mrc.ukri.org/documents/pdf/mrc-policy-on-research-misconduct/ (Accessed: 29 June 2020).

Nakamoto, S. (2009) 'Bitcoin: A Peer-to-Peer Electronic Cash System', p. 9.

Pocock, S. J. (2013) *Clinical Trials: A Practical Approach*. 1st edn. Wiley. Available at: https://www.amazon.co.uk/Clinical-Trials-Practical-Approach-Publications/dp/0471901555 (Accessed: 2 July 2020).

Sarmah, S. (2018) 'Understanding Blockchain Technology', *Computing Science and Engineering*, 8(2), pp. 23–29. doi: 10.5923/j.computer.20180802.02.

Saunders, M. N. K., Lewis, P. and Thornhill, A. (2015) *Research Methods for Business Students*. 7 edition. New York: Pearson.

Scarso Borioli, G. and Couturier, J. (2018) 'How blockchain technology can improve the outcomes of clinical trials', *British Journal of Healthcare Management*, 24, pp. 156–162. doi: 10.12968/bjhc.2018.24.3.156.

Segen's Medical Dictionary (2012) *data integrity*, *TheFreeDictionary.com*. Available at: https://medical-dictionary.thefreedictionary.com/data+integrity (Accessed: 24 June 2020).

Seife, C. (2015) 'Research Misconduct Identified by the US Food and Drug Administration: Out of Sight, Out of Mind, Out of the Peer-Reviewed Literature', *JAMA Internal Medicine*, 175(4), p. 567. doi: 10.1001/jamainternmed.2014.7774.

Sertkaya, A. *et al.* (2016) 'Key cost drivers of pharmaceutical clinical trials in the United States', *Clinical Trials*, 13(2), pp. 117–126. doi: 10.1177/1740774515625964.

Shamley, D. and Wright, B. (eds) (2017) *A comprehensive and practical guide to clinical trials*. London, United Kingdom; San Deigo, CA, United States: Elsevier/AP, Academic Press, an imprint of Elsevier.

Triall (2019) 'A brief introduction to clinical trials', *Insights by Triall.io*, 3 July. Available at: https://insights.triall.io/a-brief-introduction-to-clinical-trials-medical-research-and-pharmaceutical-innovation/ (Accessed: 2 July 2020).

Vujicic, D., Jagodic, D. and Ranđić, S. (2018) 'Blockchain technology, bitcoin, and Ethereum: A brief overview', in. Cacak, Serbia: Faculty of Technical Sciences University of Kragujevac, pp. 1–6. doi: 10.1109/INFOTEH.2018.8345547.

WHO (2018) *Clinical trials*, *Clinical Trials*. Available at: https://www.who.int/westernpacific/health-topics/clinical-trials (Accessed: 26 June 2020).

Wong, C. H., Siah, K. W. and Lo, A. W. (2019) 'Estimation of clinical trial success rates and related parameters', *Biostatistics*, 20(2), pp. 273–286. doi: 10.1093/biostatistics/kxx069.

Wong, D. R., Bhattacharya, S. and Butte, A. J. (2019) 'Prototype of running clinical trials in an untrustworthy environment using blockchain', *Nature Communications*. Nature Publishing Group, 10(1), pp. 1–8. doi: 10.1038/s41467-019-08874-y.

Zheng, Z. *et al.* (2017) 'An Overview of Blockchain Technology: Architecture, Consensus, and Future Trends', in *2017 IEEE International Congress on Big Data (BigData Congress)*. *2017 IEEE International Congress on Big Data (BigData Congress)*, IEEE Computer Society, pp. 557–564. doi: 10.1109/BigDataCongress.2017.85.

Zheng, Z. *et al.* (2018) 'Blockchain challenges and opportunities: A survey', *International Journal of Web and Grid Services*, 14, p. 352. doi: 10.1504/IJWGS.2018.095647.

Appendices

7.1 Appendix 1: Social Media (Facebook and LinkedIn) Text Invitation

Hi (name),

My name is Karl Cullen. I an Irish Masters student currently pursuing a Masters in Pharmaceutical Business and Technology in Griffith College, Dublin in association with the Innopharma Faculty of Pharmaceutical Sciences.

My research topic is on "The Impact of Blockchain Technology on Clinical Trial Data Integrity". My research involves gaining a deeper understanding into some of the predominant issues that contribute towards clinical data integrity within clinical trial research and how a new and emerging technology such as Blockchain could affect it. As part of my research, I must carry out quantitative and qualitative analysis. This consists of an online survey that will take approximately 10-15 minutes to complete. My research involves seeking those who are primarily involved with data management within clinical trials. I would greatly appreciate and value your input and participation.

As Blockchain is a relatively new technology within this industry it not necessary that you possess prior knowledge of the technology as part of my research is to evaluate the awareness of the technology. However, if you have do not have prior knowledge, a short un-biased description of Blockchain and its main characteristics will be included into the introduction to the survey.

If you would like to participate in this study, please let me know and I will provide a link to the survey. My email is: Karl.cullen3345@gmail.com

Kind regards,

Karl Cullen

7.2 Appendix 2: Privacy and Consent Agreement

This study is part of the fulfilment of a MSc degree in Pharmaceutical Business & Technology with Griffith College Dublin.

The main purpose of this research is to examine how a new and emerging technology known as Blockchain could impact clinical trial data integrity.

By answering the following questions, you are agreeing to participate in this research. Please note that all personal information will be kept confidential and will not be cited within the research and answers will be used purely for academic reasons only. All information collected will be treated as per GDPR regulations. If you do not wish to answer a question it is not mandatory, and you skip the question if you wish.

By participating you are agreeing that:

- I am voluntarily participating in this research

- I understand that my answers will be used for academic purposes only and no personally identifiable

  information will be shared

- I understand that participation is not mandatory, and I can exit the survey at any point I wish

- I did not and do not expect to receive any form of monetary incentive or otherwise to participate

- I can request a copy of my answered survey upon completion if I so wish

- I am free to contact the researcher with any questions I might have

- I understand that all questions asked are voluntary and I can choose not to answer a question if I so wish

I have read and understood the prior information and agree to participate in this study.

Your time and participation is greatly appreciated, if you have any future questions you can contact me at: Karl.Cullen3345@gmail.com

Sincerely,

Karl Cullen

7.3 Appendix 3: Original Text Invite (LinkedIn) for Qualitative Interview

Hi (name),

My name is Karl Cullen. I am a MSc student pursuing a Masters in Pharmaceutical Business and Technology in Griffith College, Dublin. My research project is on "The Impact of Blockchain Technology on Clinical Trial Data Integrity". My research involves gaining a deeper understanding into data management within clinical research and how a new and emerging technology such as Blockchain could affect it. As part of my research, I must carry out qualitative analysis in the form of structured interviews. My research involves those who are primarily involved with data management in clinical trials and I would like to have a short interview with you, if possible (by telephone or zoom), which will take approximately 15 minutes. My research involves awareness of this technology, so it is not necessary to have any prior knowledge about Blockchain beforehand. My research participants of interest are Clinical Data Managers and/or Clinical Research Coordinators.

Before considering participation, it should be noted that no personal identifiable information will be shared and although an audio recording will be kept for the interview, it will be used for transcription purposes and will subsequently be destroyed post transcription. Under no circumstances will your identity be revealed or shared.

I would kindly request your participation for this study. If you like to participate in this research, I would be happy to facilitate a date and time at your convenience and can make the necessary arrangements

If you have any questions please do not hesitate to reply to my email address: **Karl.cullen3345@gmail.com** or you can reply to me here on LinkedIn.

Kind regards

Karl Cullen